

1 ATGAGCCAGC CCAGGCCCCG CTACGTGGTA GACAGAGCCG CATACTCCCT  
51 TACCCCTCTTC GACGATGAGT TTGAGAAAGAA GGACCGGACA TACCCAGTGG  
101 GAGAGAAACT TCGCAATGCC TTCAGATGTT CCTCAGGCAA GATCAAAGCT  
151 GTGGTGTGTT GGCTGCTGCC TGTGCTCTCC TGCTGCCCCA AGTACAAGAT  
201 TAAAGACTAC ATCATTCCCTG ACCTGCTCGG TGACTCAGC GGGGGATCCA  
251 TCCAGGTCCC ACAAGGCATG GCATTTGCTC TGCTGGCAA CTTCTGCA  
301 GTCAATGGCC TCTACTCCCT CTTCTTCCCC CTCCGTACCT ACTTCTTCTC  
351 GGGGGGTGTT CACCAAGATGG TGCCAGGTAC CTTTGCCTT ATCAGCATCC  
401 TGGTGGTAA CATCTGCTG CAGCTGGCC CAGAGTCGAA ATTCCAGGTC  
451 TTCAACATG CCACCAATGA GAGCTATGTG GACACAGCAG CCATGGAGGC  
501 TGAGAGGCTG CACGTGTCAG CTACGCTAGC CTGCTTCACC GCCATCATCC  
551 AGATGGGTCT GGGCTTCATG CAGTTGGCT TTGTGGCAT CTACCTCTCC  
601 GAGTCCTTCA TCCGGGGCTT CATGACGGCC GCCGGCTGC AGATCCTGAT  
651 TTCGGTGCTC AAGTACATCT TCGGACTGAC CATCCCCTCC TACACAGGCC  
701 CAGGGTCCAT CGTCTTACCC TTCAATTGACA TTGCAAAAAA CCTCCCCCAC  
751 ACCAACATCG CCTCGCTCAT CTTCGCTCTC ATCAGCGGTG CCTTCTGGT  
801 GCTGGTGAAG GAGCTCAATG CTCGCTACAT GCACAAGATT CGCTTCCCCA  
851 TCCCTACAGA GATGATTGTC GTGGTGGTGG CAAACAGCTAT CTCCGGGGC  
901 TGTAAGATGC CCAAAAGTA TCACATGCG ATCGTGGGAG AAATCCAACG  
951 CGGGTTCCCC ACCCCGGTGT CGCCTGTTG CTCACAGTGG AAGGACATGA  
1001 TAGGCACAGC CTTCTCCCTA GCCATCGTGA GCTACGTAT CAACCTGGCT  
1051 ATGGGCCGA CCCTGGCCAA CAAGCACCGC TACGACGTGG ATTGAAACCA  
1101 GGAGATGATC GCTCTCGGCT GCAGCAACTT CTTTGGCTCC TTCTTTAAAA  
1151 TTCATGTCAT TTGCTGTGG CTTCTGTCA CTCTGGCTGT GGATGGAGCT  
1201 GGAGGAAAAT CCCAGGTGGC CAGCCTGTTG TTGTCCTCTGG TGGTGATGAT  
1251 CACCATGCTG GTCTGGGGA TCTATCTGTA TCCTCTCCCT AAGTCTGTG  
1301 TAGGAGCCCT GATCGCTGTC AATCTCAAGA ACTCCCTCAA GCAACTCACC  
1351 GACCCCTACT ACCTGTGGAG GAAGAGCAAG CTGGACTGTT GCATCTGGT  
1401 AGTGAGCTTC CTCTCCCT TCTTCCTCG CCTGCCCTAT GGTGTGGCAG  
1451 TGGGTGTCGC CTTCTCCGTC CTGGTCGTGG TCTTCCAGAC TCAGTTTGA  
1501 AATGGCTATG CACTGGCCCA GGTCTGGAC ACTGACATT ATGTGAATCC  
1551 CAAGACCTAT AATAGGGCCC AGGATATCCA GGGGATTAAA ATCATCACGT  
1601 ACTGCTCCCC TCTCTACTTT GCCAACCTAG AGATCTTCAG GAAAAGGTC  
1651 ATCGCCAAGA CTGTCTCCCT GCAGGAGCTG CAGCAGGACT TTGAGAATGC  
1701 GCCCCCCCAC GACCCCAACA ACAACCCAGAC CCCGGCTAAC GGCACCAAGCG  
1751 TGTCTTATAT CACCTTCAGG CCTGACAGCT CCTCACCTGC CCAGAGTGTG  
1801 CCACCAAGCCT CCGCTGAGGC CCCGGGGAG CCCAGTGACA TGCTGGCCAG  
1851 CGTCCCACCC TTGCTCACCT TCCACACCCCT CATCCTGGAC ATGAGTGGAG  
1901 TCAGCTTCGT GGACTTGATG GGCATCAAGG CCCTGGCAA GCTGAGCTCC  
1951 ACCTATGGGA AGATCGGCCT GAAGGTCTTC TTGGTGAACA TCCATGCCA  
2001 GGTGTACAAT GACATTAGCC ATGGAGGCCT CTTTGAGGAT GGGAGTCTAG  
2051 AATGCAAGCA CGTCTTCCC AGCATACTG ACCGAGTCCT CTTTGGCCAG  
2101 GCAAATGCTA GAGACGTGAC CCCAGGACAC AACTTCAAG GGGCTCCAGG  
2151 GGATGCTGAG CTCTCCCTGT ACGACTCAGA GGAGGACATT CGCAGCTACT  
2201 GGGACTTAGA GCAGGAGATG TTGGGAGCA TGTTTACGC AGAGACCTG  
2251 ACCGCCCTGT GA (SEQ ID NO:1)

**FEATURES:**

Start Codon: 1  
Stop Codon: 2260

**FIGURE 1A**

Docket No.: CL000861CON  
Serial No.: TO BE ASSIGNED  
Inventors: Karl GUEGLER et al.  
Title: ISOLATED HUMAN TRANSPORTER...

**HOMOLOGOUS PROTEINS:**

**Top BLAST Hits:**

|                          |   | Score | E     |
|--------------------------|---|-------|-------|
| gb AAF81911.1 AF279265_1 | (AF279265) putative anion transpor...         | 476   | e-133 |
| gb AAF71715.1 AF230376_1 | (AF230376) prestin [Meriones ungu...          | 471   | e-131 |
| ref NP_000432.1          | pendrin [Homo sapiens] >gi 11421915 ref XP... | 451   | e-125 |
| ref NP_035997.1          | Pendred syndrome homolog (human); Pendred'... | 448   | e-124 |
| ref NP_062087.1          | Pendred syndrome homolog (human) [Rattus n... | 447   | e-124 |
| ref NP_067328.1          | down-regulated in adenoma [Mus musculus] >... | 434   | e-120 |
| ref NP_000102.1          | down-regulated in adenoma protein [Homo sa... | 418   | e-115 |
| sp 070531 DTD_RAT        | SULFATE TRANSPORTER (DIASTROPHIC DYSPLASI...  | 365   | 1e-99 |
| ref NP_000103.1          | sulfate anion transporter 1; Diastrophic d... | 362   | 1e-98 |
| ref NP_031911.1          | diastrophic dysplasia [Mus musculus] >gi 2... | 357   | 4e-97 |

**BLAST to dbEST:**

|   | Score | E     |
|---|-------|-------|
| gi 8630793 /dataset=dbest /taxon=960... | 523   | e-146 |

**EXPRESSION INFORMATION FOR MODULATORY USE:**

**library source:**

**Expression information from BLAST dbEST hits:**

gi|8630793 Human head-neck

**Expression information from PCR-based tissue screening panels:**

Human fetal lung

**FIGURE 1B**

1 MSQPRPRYVV DRAAYSLTLF DDEFEKKDRT YPVGEKLRNA FRCSSAKIKA  
51 VVFGLLPVLS WLPKYKIKDY IIPDLLGGLS GGSIQVPQGM AFALLANLPA  
101 VNGLYSSFFP LLTYFFLGGV HQMVPGTFAV ISILVGNICL QLAPESKFQV  
151 FNNATNESYV DTAAMEAERL HVSATLACLT AIIQMGLGFM QFGFVAIYLS  
201 ESFIRGFMTA AGLQILISVL KYIFGLTIPS YTGPGSIVFT FIDICKNLPH  
251 TNIAISLIFAL ISGAFLVLVK ELNARYMHKI RFPPIPTEMIV VVATAISGG  
301 CKMPKKYHMQ IVGEIQRGFP TPVSPVVSQW KDMIGTAFSL AIVSYVINLA  
351 MGRTLANKHG YDVDSNQEMI ALGCSNFFGS FFKIHVICCA LSVTLAVDGA  
401 GGKSQVASCL VSLVVMITML VLGIYLYPLP KSVLGALIAV NLKNSLKQLT  
451 DPYYLWRKSK LDCCIWWVVS LSSFFLSLPY GVAVGVAFSV LVVVFQTQFR  
501 NGYALAAQVMD TDIYVNPKTY NRAQDIQGIK IITYCSPLYF ANSEIFRQKV  
551 IAKTVSLQEL QQDFENAPPT DPNNNQTPAN GTSVSYITFS PDSSSPAQSE  
601 PPASAEAPGE PSDMLASVPP FVTFHTLILD MSGVSFVDSL GIKALAKLSS  
651 TYGKIGVKVF LVNIHAQVYN DISHGGVFED GSLECKHVFP SIHDAVLFAQ  
701 ANARDVTPGH NFQGAPGDAE LSLYDSEEDI RSYWDLEQEM FGSMFHAETL  
751 TAL (SEQ ID NO:2)

**FEATURES:**

**Functional domains and key regions:**

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATION  
N-glycosylation site

Number of matches: 3

|   |         |      |
|---|---------|------|
| 1 | 153-156 | NATN |
| 2 | 156-159 | NESY |
| 3 | 580-583 | NGTS |

[2] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
Protein kinase C phosphorylation site

Number of matches: 2

|   |         |     |
|---|---------|-----|
| 1 | 45-47   | SAK |
| 2 | 445-447 | SLK |

[3] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
Casein kinase II phosphorylation site

Number of matches: 11

|    |         |      |
|----|---------|------|
| 1  | 18-21   | TLFD |
| 2  | 158-161 | SYVD |
| 3  | 240-243 | TFID |
| 4  | 365-368 | SNQE |
| 5  | 459-462 | SKLD |
| 6  | 556-559 | SLQE |
| 7  | 635-638 | SFVD |
| 8  | 691-694 | SIHD |
| 9  | 722-725 | SLYD |
| 10 | 726-729 | SEED |
| 11 | 732-735 | SYWD |

[4] PDOC00007 PS00007 TYR\_PHOSPHO\_SITE  
Tyrosine kinase phosphorylation site

Number of matches: 2

|   |         |           |
|---|---------|-----------|
| 1 | 7-15    | RYVVDRAAY |
| 2 | 447-454 | KQLTDPYY  |

**FIGURE 2A**

[5] PDOC00008 PS00008 MYRISTYL  
N-myristoylation site

Number of matches: 10

|    |         |        |
|----|---------|--------|
| 1  | 77-82   | GGLSGG |
| 2  | 78-83   | GLSGGS |
| 3  | 89-94   | GMAFAL |
| 4  | 103-108 | GLYSSF |
| 5  | 335-340 | GTAFL  |
| 6  | 435-440 | GALIAV |
| 7  | 481-486 | GVAVGV |
| 8  | 485-490 | GVAFSV |
| 9  | 581-586 | GTSVSY |
| 10 | 681-686 | GSLECK |

**Membrane spanning structure and domains:**

| Helix | Begin | End | Score | Certainty |
|-------|-------|-----|-------|-----------|
| 1     | 51    | 71  | 0.893 | Putative  |
| 2     | 82    | 102 | 1.020 | Certain   |
| 3     | 107   | 127 | 1.729 | Certain   |
| 4     | 130   | 150 | 1.497 | Certain   |
| 5     | 186   | 206 | 1.723 | Certain   |
| 6     | 228   | 248 | 1.517 | Certain   |
| 7     | 256   | 276 | 1.898 | Certain   |
| 8     | 288   | 308 | 1.252 | Certain   |
| 9     | 338   | 358 | 1.568 | Certain   |
| 10    | 383   | 403 | 1.304 | Certain   |
| 11    | 412   | 432 | 2.345 | Certain   |
| 12    | 469   | 489 | 1.997 | Certain   |
| 13    | 619   | 639 | 1.146 | Certain   |

**FIGURE 2B**

**BLAST Alignment to Top Hit:**

>gb|AAF81911.1|AF279265\_1 (AF279265) putative anion transporter 1 [Homo sapiens]

Length = 738

Score = 476 bits (1224), Expect = e-133  
 Identities = 263/724 (36%), Positives = 428/724 (58%), Gaps = 36/724 (4%)  
 Frame = +3

Query: 54 LFDDEFEKDR--TYPVGEKLRNAFRCSSAKIKAVVFGLPVLSWLPKYKIKDYIIPDLL 227  
 Sbjct: 15 L + EE R + P + R +CS A+ A++ LPVL WLP+Y ++D++ DLL 74

Query: 228 GGLSGGSIQVPPQGMAFALLANLPAVNGLYSSFFPLLTYYFFLGGVHQMVPGTFAVISILVG 407  
 Sbjct: 75 GLS +Q+PQG+A+ALLA LP V GLYSSF+P+ YF G + GTFAV+S++VG 134

Query: 408 NICLQLAPESKFQVFNNATNESYVDTAAMEAERLHSATLACLTAAIQMGLGFMQFGFVA 587  
 Sbjct: 135 ++ LAP+ A N+S ++ A +A R+ V++TL+ L + Q+GLG + FGIV 186

Query: 588 IYLSESFIRGFMTAAGLQILISVLKYIFGLTIPSYTGPSSIVFTFIDICKNLPHNTIASL 767  
 Sbjct: 187 YLSE +RG+ TAA +Q+ +S LKY+FGL + S++GP S++T ++C LP + + ++ TYLSEPLVRGYTTAAAVQVFSQLKYVFGHLSSHSGPLSLIYTLEVCKLPQSKVGT 246

Query: 768 IFALISGAFLVVKELNARYMHKIRFPIPTEMIVVVVATAISGGCKMPKKYHMQIVGEIQ 947  
 Sbjct: 247 + A ++G LV+VK LN + ++ PIP E++ ++ AT IS G + ++ + +VG I VTAAVAGVVVLLVKKLNDKLQQQLPMPPIPGEELLIGATGISYGMGLKHRFEVDVGNIP 306

Query: 948 RGFPTPVSPVVLQWKDMIGTAFSLAIVSYVINLAMGRTLANKHGYDVDSNQEMIALGCSN 1127  
 Sbjct: 307 G PV+P + ++G+AF+A+V + I +++G+ A +HGY VDSNQE++ALG SN AGLVPPVAPNTQLFSKLVGSFTIAVVGFAIAISLGKIFALRHGYRVDNSQELVALGLSN 366

Query: 1128 FFGSFFKIHVICCALSVTLAVDGAGGKSQVASLCVSLVMITMLVLGIYLYPLPKSVLGA 1307  
 Sbjct: 367 G F+ + C+S +L + GG SQA SL +++ ++ LG + LPK+VL A LIGGIFQCFPVSCSMSRSLSVQESTGGNSQVAGAISSLFILLIIVKLGEFLHDLPKAVLAA 426

Query: 1308 LIAVNLSKNSLQLTDPPYYLWRKSKLDCCIIVVSFLSSFFLSPYGVAVGVAFSVLVVVFQ 1487  
 Sbjct: 427 +I VNLK L+QL+D LW+ ++ D IW+V+F ++ L+L G+ V V FS+L+VV + IIIVNLKGMLRQLSDMRSLSWKRANRADLLIWLVTFTATILLNLDLGLVVAIFSLLLVVVR 486

Query: 1488 TQFRNGYALAQVMDDTDIYVNPKTYNRAQDIQGKIKIITYCSPLYFANSEIF----- 1637  
 Sbjct: 487 TQ + L QV DTDIY + Y+ A++++G+K+ + +YFAN+E + TQMPHYSVLGQVPDTDIYRDVAEYSEAKEVRGVKVFRSSATVYFANAEIFYSDALKQRCGV 546

Query: 1638 -----RQKVIAK--TVSLQELQQDFE-NAPPTDPNNNQTPAN-GTSVSYI----- 1760  
 Sbjct: 547 ++K++ K + L++LQ++ + P N TS+ + DVDFLISQKKLLKKQEQQLKLKQLQKEEKLRKQAASPKGASVSINVNTSLEDMRSNNVED 606

Query: 1761 -----TFSPDSSSPAQSEPPASAEAPGEPSDMLASVPPFVTFTLILMSGVSFVDLMGI 1925  
 Sbjct: 607 S D A + ++AP + S + A P FH+LILD+ +SFVD + + CKMMQVSSGDKMEDIATANGQEDSKAP-DGSTLKALGLPQPDFHSLILDLGALSFVDTVCL 665

Query: 1926 KALAKLSSSTYKGIGVKVFLVNIHAQVYNDISHGGVFEDGSLECKHVFPsiHDAVLFAQAN 2105  
 Sbjct: 666 K+L + + +I V+V++ H+ V + + G F D S+ KH+F S+HDAV FA + KSLKNIFHDFREIEVEVYMAACHSPVVSQLEAGHFF-DASITKKHILFASVHDATFALQH 724

Query: 2106 ARDV 2117  
 R V  
 Sbjct: 725 PRPV 728 (SEQ ID NO :4)

**Hmmr search results (Pfam):**

| Model   | Description                              | Score | E-value | N |
|---------|--|-------|---------|---|
| PF00916 | Sulfate transporter family               | 254.5 | 1.5e-72 | 1 |
| PF00189 | Ribosomal protein S3, C-terminal domain. | 3.3   | 8       | 1 |

## FIGURE 2C

Docket No.: CL000861CON  
Serial No.: TO BE ASSIGNED  
Inventors: Karl GUEGLER et al.  
Title: ISOLATED HUMAN TRANSPORTER...

Parsed for domains:

| Model   | Domain | seq-f | seq-t | hmm-f | hmm-t | score | E-value |
|---------|--------|-------|-------|-------|-------|-------|---------|
| PF00916 | 1/1    | 187   | 497   | ..    | 1     | 328   | 1.5e-72 |
| PF00189 | 1/1    | 651   | 661   | ..    | 79    | 89    | 3.3     |

**FIGURE 2D**

1 CTGGGTTCT ATGTGGGGAG GTCATGCTCC CCACTCATTG AGCCCCCCC  
51 GGCACAAACAC CTGGACAGGC AGACCCATGC AGACTCTGGA GCAGGTGGAG  
101 AGGAAGAGTG AGACCACCCC GCCTCACGGG CGGTGAAGGG CCGGCAGCCT  
151 CTGAATAGTC TCTGCTAGGA GGTAGAAAAGC ACCCTCCCAT CTTAATCATA  
201 GTAATCATCG CCACTACCAT TTACTGGGTG CCTATAAAAG GCCAGCCTCT  
251 TCATAACACAT GATCTCACTG AATCCTCATA GCATCTGCCT GCGACTGTTA  
301 TTATCCCCAT TTACAGATGA AGAAACTGAA TCTTGAACC CAGGTCACT  
351 GGCTCTCAA CTTGTGCTGT TTTCCCTAAG CCACCCGGTC TCTCATTCT  
401 CCCACTGAAA TGTCTCACAT GCCATTGCC TTACTCATTT CTGCCCAGT  
451 CTCCCTCCAA ACACCATTTA TCAATTGCCT CAACAAGTAT GTGTTGAGTA  
501 CACACTAAGG GCCAGGGCAGG GGGCTGGCA CAGGCCTGCGGG GGGTAGGTT  
551 ATTCTCCAC CTTCGCTCT GCTGGGTATC ACCTGTGGGG TCTTGCCGGG  
601 CATCCCACCC TCACCTGTAG TTCAAGTGGA CCTTGGGATC CCAAGACCAA  
651 ATGAATGGAA TGCACCAGCC CAGCCTTCAC CAACTTGAGC ACAATCTTAT  
701 TCATAATAGA AACTCACATT TGCACTCACAC TTTACATTTT ACACAACCCC  
751 TTCTTATCCA TTAACTCATT TGATCTTCAC AACAACCTG TGAGATATGT  
801 CTGTTACTCC CACTTTAGTG ATACAGAAC TGAGGTTTGAAAGTAATG  
851 TGACCATTTCT GCCTCATTAA TAAAAGCAGG ATTAACCCAG GCTCCTGGAC  
901 CCTTCCACAA AAGGCATTAA GCAACCTGCT CCCCTCTGAC AACCTCCCT  
951 GTCACCCAGG CTCTCCCTG GGAAGTTGGG GGCATCTCA GCCCCCAAGT  
1001 AGTTACTCAT TTTCAACCCCC ATCTCAAATC TTTTGCCAAA CTGGCCACAG  
1051 CCACCCACAA CTCCCCACCT CCCAGATACA AATCCTCACT CTAAGCCTTC  
1101 CCCATCTCTT CTTCTCTGT CCTTCTTCT CTGTGGTCCT CTGAGCAACT  
1151 TCTCCCAAGCT CTGGGAGGT GAGGGGAGGT GGGAGACCCA GTAAATTGAA  
1201 GAGGGAGGGG GAAAGGTTCC TACAGGGAAC TCTCCGGGG CTCAGGGGCC  
1251 CTGGCACTCA GCTCTGCCCA TCTCAGCTCC TGGAACGTCA GCCAGGTTGC  
1301 GCAAAAAGTG AGGAGGAGAG GAGCGGCAGT ACACAAGGGT GGGGGAAAGA  
1351 TTAGGCACAG GAAGCCGTGG GAGAGAGAGC CGGCAGGTGG ACCATCCTGG  
1401 TTTCCCCACA CACACCATTG TCCCCCTGGG AAACCTGTTG GTGAAGTTCT  
1451 AGATGTCTT TCCAAGAAGG GTCTCTTGA GTCTCTTCA GCTATCCCC  
1501 TGCCCTTAGG CAAGCTGTTT TCTGTTTCTT CCAAGCTGAC TGGCTGAATG  
1551 GTAGGAGCCT TTCTGCCAGG GAAACTAAGG TCTGGGAAGG GAGTATGGCT  
1601 TGTGGGGACA CCAAGGGTCA GGGGAGGGGA GGGTCCACCT GCTGAATCAA  
1651 GTGGGGCTC CTGCCCTCGT GATTCCCCCT TGCCTGGTGC TCAGTGGGG  
1701 TGATGGTGAC GCCACAGGTG TGGAGTGCCA GCCACGTGCT GAGCGCCAAG  
1751 CAAAACGCC AGGGTGTAGTC TATGCATCAT CAGTGCCTGG GAAGGAAGGC  
1801 CACTGGCAGG AGGGAGTCTG ACGGAAAAAC TTGACAGAGG GAAGGGAGGC  
1851 ACCTTGCTT ATCGGGCGGG GGAAGGCCAG AATAAAACTC TGACTTGCA  
1901 AGGACCAAGAG AGAGAAGGCC TGGGCTGGCA CTAGGGAGGG ATGTTCCCTC  
1951 ACCCTCCCT CCTCTGCTTC TCCCAAAGCT TGAAATGCC CCAGATATGA  
2001 GCCAGCCCAG GCCCCGCTAC GTGGTAGACA GAGCCGCATA CTCCCTTACC  
2051 CTCTTGACG ATGAGTTGA GAAGAAGGAC CGACATACC CAGTGGGAGA  
2101 GAAACTTCGC ATAGCCTTCAG GGTAACTGTT CCAGAGGCCA GACTTCTGCC  
2151 TCCTCTGCTC CCTACAAAAA TCCTTCTGC ACCAGGACAC GGCTTCTGCA  
2201 CTGGTATCCC TAAGATGGGG TTAAGGGAGG CCTCTGGGGAA GTGAGGTTCT  
2251 GAATGATGAA TTAAAGATCC TACAACCTCA TCTGTACTGA GACCCCCCAGG  
2301 GAGGATGGGG AGCAGGAGCA AGAACCATCC AGAAGGGTTA TATGGCATT  
2351 CCAAACCCCT GCATGGCATC TCCCATATTCA TCAATTCACT CGGGTCTCTC  
2401 TGGGTTTGT AAGGCATGGT AGATGAGCAT CTACGTTATG GAGGGGTGGG  
2451 GAGCATCAGA GCCCTTACTC CATGCCCTGT TCCCTCCCTTA CAAAAAATAC  
2501 CTGAAGTTAC CATCACCCCCA GGTCTTTGT CCTTCTCCCTC CGGGATGTT  
2551 CTTCTCCAC TTGGTCCAGA GAATGCCAAA AGGAGGCCCT AAATTCTGAA  
2601 ACTTTCTGA GGGGACCTAC CAGGGTGTAG TCTTACCAAGC GCCCAGGGTC  
2651 TTTCACTCT CATCTCCCTG GAAATGCGAT GGTGGGTATG AAACCTGTC  
2701 CCTAAGTAGG CGCTACACAA GGTGATCCAT ACCCACACCC CAGGAGGCTG  
2751 GGGCTGCGGG TGTCACCCCTC CCCATTCCCA GACTCCTGGC AGACCTCCTC  
2801 TGGCCCAAGCT ATAGGCCAAC TCACTCTCCC TCACTCCCTT GGGGAAACGG  
2851 CTGATTCACT TACCTGGATT GAGGTCACTG GCAATGGCTG AAGTGGAGAC  
2901 GCAGGGTGGAA CTGGTTCAAGG CCGGGGGAAAT CACCCACTTG AGTTTGACT  
2951 AAAAGCCCCA GCCCAGCCCT GTTTCTCTTGGAGGCTCCA TTTCTGCCCA  
3001 GTTACAGTCT GTCTCACAG CTGTGCTCT CAGACAGGTG GTCTCTGCCA  
3051 GTCTTGTGC CCAAGACTTT AGGGCACAAA GTCTGAGGAT GAGAAGATCT  
3101 GCTATTGTC TAAAAGATTA GGATAATGAA AGCTGTAAAG GGATATAGCA  
3151 AACTAACAAAT CCTCTATGATA CTGGCATGAG AGCCTTGAAC AGTGCCTGGC  
3201 ATAGAAGG TGACCAATA AATATTGTT TCAATGAAAT ATGAATGAAT  
3251 GAATGTCTAG AAAGCTAATC CCTCTCAGCC TCTGTTTCA GTTCTCTT  
3301 CAAGCTTCAG ATTGCTTTGC CCAACATACA GCAGACTTGC AAGTAAGGTT  
3351 GGGCATGGAC TAGCCCTCAA ATGAGTTGTT TTTCTTCCC TAGCCAGCTC

## FIGURE 3A

3401 TCTATTCTATA AGTCCGGCTT TCTCTGCCAC AAACAGACCT GATGGAGCCC  
3451 CTGCAGGGCT GGTTCCTCTC TCAAGCAAGG CTTTAGAGTT GCATTAAGCA  
3501 ATTTATCCCC CGTCCACCTC CCCTTCAGC ATCCCAGGGA TGGCAGAGGC  
3551 ACCCATGAGC CCCAGAAGGG ACAGGGGTA AGATATTGAT GATGATGCTT  
3601 TTTCTTGGAG TGTTAGTTGG AAGAGAAAAT CTGCCCAGAC TTTCCAAGGT  
3651 ACAAAAGCATT GTCTTGTGTT GTTCAGTCT TGTTGACAT CCAGGGGACCC  
3701 GAGTGTCAAGG GAAACTATTG TTGAGCAAGA GCAAAGAGCA GGAATTGGTG  
3751 CTGGCAGGA AAGGAAGCCT CATCAGAGCA GGGCAGTGAG TCACCAAATG  
3801 GCCCTTAAGT ATTTGAGTTC CCTCAACTGG GAGAAGGAAA GCAAATGCC  
3851 CTCAACCACT CCTAGTCATC AATCCACCGG CTGTCACCCCT TGAGTTGTA  
3901 AGCCCTGTT CCTACCGCTC CTGAGTTCT ATGAAAGGAC CTTGAGGTGT  
3951 TCAACAAACA GGGAAAGGGAT CAACTCTCCC CACCCCTGCGT TGACCAATGA  
4001 ATTCTTCCCT CCTCTGCTGC CCAGTGAATT AACAGGAGAA AGAACTCCGG  
4051 TATGGAGTT ACCACACATA AAGGATAGTG AGTCAGCAGA GTGCACCTG  
4101 CAGGAAACAT AGAGCCTTCC TTTCAAGGA ATTTCTAAGA AAAATGGCAG  
4151 CAGGCAGGGC CCAACTGGGT GTATTCACTC ATTCACTTTAT TCAACAAATA  
4201 TTACTAAGT GCCCCCTGTC AAGGCTCGAG GTGTACAAAG ATGAACAGGA  
4251 GAGCTAGACT TCTTGCCATG CGTGGTGGGG TTTGCTGCCT AGTGGGAGAG  
4301 ACAGACAAAA AGCAAGGAAT GCACACACAG GATGCACACA CAGCGGAGG  
4351 AACCAAGGTG CAGTTACCCA GGCCTGGAT CAGACAGACAGA GGACTCAGAG  
4401 GAGACTTCC CAGAGAAAAG CCATCTGAGC CAAGGGATGG ATCTGATACC  
4451 TCCGAAGGCT GAGCCACCAT AACACTCATC CCTTTAAGCC AAGTCTTATA  
4501 AACTCCCCAG GTAAAGCAGT GGCAGTCAGA AGACCTCCAG CTAATGCCCA  
4551 GGACAAGTTG ATGAGCTCTC AAGAAAAAGT TCTTGCCCTT TCTTCTCAAT  
4601 ATCCCTGGCA CACAGTTCA TGATTGTTGA ATGAACCAAT GAATGAAATG  
4651 AGCAGGATAT GATAATCCCT CTCCAACACG GAATGTCCAA GCCATGCAGA  
4701 GCCGACTTGG AATTTTCCCC GTTCCCTTCC AGATGTTCCCT CAGCCAAGAT  
4751 CAAAGCTGTG GTGTTGGGC TGCTGCCGTG TGCTCTCTGG CTCCCCAAGT  
4801 ACAAGATTAA AGACTACATC ATTCTTGACCC TGCTCGGTGG ACTCAGCGGG  
4851 GGATCCATCC AGGTCCCAAGA AGGTGAAGGG GCTCCTTCAG CCAGGCTGG  
4901 ATTGCCACTC CCCTCACCAT TCCCTCTCTC ATCCCCACTC CATCCCTCTG  
4951 TGATCCCCAT AAGCTAGTCA TGCTGCTGAG CTTCACTC TGTTGCTCT  
5001 GCAGGCATGG CATTGCTCT GCTGGCAAC CTTCCTGCAG TCAATGCC  
5051 CTACTCCCTC TTCTTCCCCC TCCCTGACCTA CTTCCTCCTG GGGGGTGTTC  
5101 ACCAGATGGT GCCAGGTAAG GCCTCTCCC TCTGGGCAGG CAGGATGACC  
5151 CAGACACAAAG GGATGGGAGG TGTGGCAAG GGGCCTCGGG AGATTTCCA  
5201 TCTGCATTCT CTTGGAGTTG TTCTGGTCA GTCTCTAGGGG AATGGTCACT  
5251 GTGAATGTCA TTTCCAGGTC CTCGGTGACC TTGGAGAAAC CACTGAGCCT  
5301 CTTTGAGTTG AGTTAGCATT ACCTGTTCCA TCTTCCTCCT AGGAATGAGA  
5351 GGAAGACTTA GCAGAACAAAG ATATACCATA TGCTATAACA TGCTTAAACA  
5401 GATGTGAGAA ATCACCACAT AACTCCCTGG TTGGTCCCAG CGGCCACTA  
5451 CAGGGACATCCTTGGGACTCTC TGTTGCTAAG TGAGATGGAG GAAAGCCTGG  
5501 TCACAGGGC TTGTTTCTGG TTCAAGCTCT GCTTATATTCTT CTTATTCTG  
5551 AGTCATTTT CTCACGTGTC CTGTATGACA ATATTGACCA TTGGGGTAAA  
5601 AGCACCTTGA AAAGCATAGA TCATGGTTAG AGTGAGTGTT TGTTATTATT  
5651 GTGTTGGAGA AGAGCCTTGG AGGTGCAGGG ATCCATCCCC CTGGGGTCTG  
5701 GAAGCATTCC TGGGGCCCTT TCTGGTTTCC ATCGGTGTGG TTCAAACCTC  
5751 TGATTTTGC TGGCTGGGT GGGCACACCA GGTACCTTTG CCGTTATCAG  
5801 CATCTGGTG GGTAAACATCT GTCTGAGCT GGGCCCAGAG TCGAAATTCC  
5851 AGGTCTCAA CAATGCCACCC AATGAGAGCT ATGTGGACAC AGCAGCCATG  
5901 GAGGTGAGA GGCTGCACGT GTCAGCTACG CTAGCCTGCC TCACTGCCAT  
5951 CATCCAGGTG AGGGGGCAGC CCCCCAACCT GCTAGAAGGG CATCAGACCA  
6001 CCCTGCCCTT CCCTCAAAGC CTTAGTTTG ATGCTAAATC TGATTTAGGG  
6051 GGCTGGGTGT GGAGGCTCAT GCCTGTAACT CCAGCACTTT GGGAGGCTGA  
6101 GGAGGGTGGAGA TCACTTGGAG TCAAGGAGTTT GAGACCACT TGACCAACGT  
6151 GATGAAACCC CATCTCTACCA AAAAATACAA AAAATATCCA GGCTTGGTAG  
6201 TATGCGCTG TAGTCCCACCC TACTCAGGAG GCTGAGGAG GAGAATCACT  
6251 TGAATCCGGG AGGCAGAGGT TGCACTGAGC TGAGATCGCG CCACTGCACT  
6301 CCAGCCTGGG TGACAGAGGC AGACTCCGTC TCAAAAAAAA AAAAAAAA  
6351 AAAAAAAA CCAAGTTAG GGCTCACCTC CTCCCTCCTC CCCATCCCAG  
6401 GGCTAAAGTG AACCTTGAAA ATTAACAGTA TCTCTCTCATC TGCTGTAGC  
6451 AGGACCATAC AAAAAAAACAA CAGCTGTACCC TGTTAAACT GCTCTGAGCT  
6501 TAAACCTGT AAAAGACTCA CAGCCTCTCT CCATTATCCC GTGGAGAAAC  
6551 CCAACTCTCT GCCAGCATAG TCTTGAGAC TGCTAATTTT CTCTAACATC  
6601 CCTCACTCCG CTCCAGCCTC CTCTGCTCCA AGCCACAGCA GCAGTTGCAC  
6651 AACATAAAATT GAGCTTCTGC AAATGGTTGC AAAGGATTCT GCTAGTTTT  
6701 ATGAAGGGAA GCACAAACATG ACAGAAATGCA AGAGCAAAAC ACAGTCCAG  
6751 AGAGCGCTT TTCACTTCACT CATTCTTCAG GTTTTGTGCC AAGAACTAGG

FIGURE 3B

6801 CTAAACCTG GGATACAAAG ATAAGTAAGA AAGAGGTCCA ATTACAAGT  
6851 TGCTCACAGC CCAGCAGG AAGGAGCCAT GTCAACAGAT AAATTTGTAT  
6901 GCAGTGAGAT AACAGCAAA GTAGAGCCAT GTCAAAGAC TGTAGGGACA  
6951 CAGAGCAGAG TCACGGAGA CCTCAAAGAG GAGGTGACAC TCCACCTCTC  
7001 TAAAGGATG AGAACTTAAC CAGGAACAAG GTATAACAGAG GATGGTCAG  
7051 GCAGAAGGGA ACAGTGCCTA AAAACACTGA GGCCTGAGAG AGTGTGATCT  
7101 GCGCAGGCAA AGTAAGGGGC TTGGTGTGGC TGAGGGTAG AGGGCCAGA  
7151 AGAGGATGGA AAAGTAGGCA GGAGCCAGAC AATGAGATCT GGGGTCTGTT  
7201 CTCTGACAGC GACTTTGGG CTGATTGGC GTTTATAAGG ATCGTTTGGG  
7251 CTACAACTG ATGAGTGGG GTGGGATTAG AATCAAGGCA GGGGACCTGT  
7301 TGGGAGACTC TGCAAGGGCC CAGGCAAGAA TAATGCAGGC GAAGACCAGG  
7351 TAGAGAAAGA GATGGGGCTG GACTGAAAAA GAATGTTTA CCAGGAGCTT  
7401 GGTGATAGAC TGGATGTGGG AGGTAAAGGA GGATGACTCT CAAGTTTTG  
7451 GTTGGGCAAC CAGGTTAATG ATGGTGTAT TTACTGAGAG AGAAAACACT  
7501 GGGGAGGAC TAGACTTTT TTACAGATAA GCAAAGCCA GAGGGTGTAT  
7551 GTGACAGAAA GGGCCATGCT CTAAAGGAGC TGAAGGTCTG ATGGCAGCCA  
7601 TGTAGAGCAC AGTGAAGGGC AGGTGAAGGT CACAGATGGT CCAATTCCCT  
7651 CAAGCTACTG CTACGCTAGG ACTGCAACGG GCTCCAGACC TCGTGTGTT  
7701 TGGGGCGGGT CGTTGGAAC GCTGAACAC ATTGGTCTTC CGCCACCAAC  
7751 CACCTTTTC CTCCCTCTAG ATGGGCTGG GCTTCATGCA GTTGGCTTT  
7801 GTGGCCATCT ACCTCTCCGA GTCTTCATC CGGGGCTTC TGACGGCCGC  
7851 CGGCCTGCAG ATCTGATTG CGGTGCTCAA GTACATCTTC GGACTGACCA  
7901 TCCCCCTCTA CACAGGGCCA GGTCATCG TCTTGTGAG TCTGGGGATG  
7951 CACCCCTGCC ATTGGAGCAA GGCTCCAGCA GACACATGAG GAGGATGTAC  
8001 TGTTTAAGA TGTCTGAGC TCCTCATTCG AAGGGCTGGC TTAGCTGTTG  
8051 TTCAGAGAGG ATTCTGAGGG GTTCTGTC TTGGGAGGGT CAAAGTCATG  
8101 ACTCACAGAG GTTCTTGGT GTAAATACCT GCAGAAAAGA GCTGTACATT  
8151 CTCCGCCAGT TCCCCATTCT AGTGCCTCAA CCCTCCCTG CCTGGAAGT  
8201 CCTGGCTTAT GTCTAATCTC CATCCCCCTC CCTTCAGGCC AAACCTTTCT  
8251 AAAGAAAAAG AAAGCATTCC TTTTCTAGCA CAAGTTCCCCC ATGTGCTTT  
8301 TGGGAAAGGG CGGTGGGCGA CGGGACAGGG TTCTGATCA GGGTTTTAAT  
8351 TCTGTCTGG TGTGCTCCA TTAGCTTGA TGGCATCCCT TCCCTGGTC  
8401 AGACACCCAA AGGTGGGGTA TTATGGAAG AAGGGGTGGG AGCCTGTGAG  
8451 CATGATGCTC TTTCCCCCAG ACCTTCATTG ACATTTGCAA AAACCTCCCC  
8501 CACACCAACA TCGCCTCGCT CATCTTCGCT CTACATCAGCG GTGCCTTCT  
8551 GGTGTGGTG AAGGAGCTCA ATGCTCGCTA CATGCACAAG ATTGCTTCC  
8601 CCATCCCTAC AGAGATGATT GTGGAAGGA CCTTGTTCAG AGCTGGGATG  
8651 TTGGGGGCC AGGCTGTGAG ACAGAGGAAGC CCTTACCTT CCTCACCCCA  
8701 TCCCCCTAAC TGGCAGCCAG TGGGACAGGA AGTCAGTTGT GAATCCATCC  
8751 CATCCCCCGT ATGTGGCTT TCCTCTCTT CTACTGCTCT AATAATTCCC  
8801 CCTAAGGAGG CAGGGAGGTG GGATTCAAGG TCCCCAGAGA AAAGGGAGAC  
8851 TTGAGAGAGA CGCCTGCCCT GGCCCCCACCT TAGGGCCAAT CCCCATTCTC  
8901 CACTCTGGG TTGCAAGGTG GTGGTGGCA CAGCTATCTC CGGGGGCTGT  
8951 AAGATGCCA AAAAGTATCA CATGCAGATC GTGGGAGAAA TCCAACGCCG  
9001 GTGAGTCCAG GTGGCCCAGA AGCCTGGCC ACCCGCACCT CATCCCCAC  
9051 TAAGCCTGA GCTCGGAGAG GGAGACAAGA TGAACCTTAT GAAAGTGCAG  
9101 TCGAAACTGT ATGACACTGA CCATGTATGA ATTATTACTA TTACCGTTT  
9151 CTGAGAAGGG CGCACAACC AGCCAATGTA GGCTATTTTA TGAGAAATGA  
9201 GTCTTAACCTG CCACACTCCC CTATAAAATC TCATTCAACT GATGCTGTTA  
9251 AACAAAGCCT CTCTGAACAG CCGCTTGCTG GCTTTTGCC TTGCTTAAT  
9301 GCATTGGTTC TTTGTCCATG TAGAAAGGA ACTATTAGGT TCAACCAGAT  
9351 TCATGAAGCA TCCACTCTGT GCCAGGCACC ATGCTGGGCC CTGGGAGGAG  
9401 AGGGGTGACG CTTGCTCTGC AGGGTTGGAA CAGGCAAGGG AGGGAAGACC  
9451 ACATAGCACC AAAGGCTCTAG GGGTCTGTGG ACTCGTGAGC ATACAGGGTT  
9501 CAGAATCTGG GAGTTAACAA ACGAGGCCCT ACCACATACT GGGCCGGGAA  
9551 CCTGGGCCAA GTTAGGTTCT CTCAGCTCA GTTTCCTCTT TTGAAAACA  
9601 GGAGTGATGG TCCCTACCCCT ATGGGGTGGT GCTGAGGATT CAGACTGGAT  
9651 GGGATAACTT AGGCAAAGAT CCCGGCACAC CATGGGGGCC TGGCTGGTCC  
9701 CTGTGGGCTG GTGAAGGACT TGGCTGCCCT CCCCCACTCAC ACCCTTGGGT  
9751 TCTGCCTCTT TCCTGGCTCC TCGGCAGGT CCCCACCCCG GTGTCGCTG  
9801 TGGCTCACA GTGGAAGGAC ATGATAGGCA CAGCCTCTC CCTAGCCATC  
9851 GTGAGCTACG TCATCAACCT GGCTATGGC CGGACCCCTGG CCAACAAAGCA  
9901 CGGCTACGAC GTGGATTGCA ACCAGGTAGC TCTGGCCACC CCCGGCAGGA  
9951 CTGGGCAGGA CAGGTCAACT CAGGCCTGGC ATGACATATC TTGGGTGGGG  
10001 AGATCATTGG GCTGAGGTTGA GGCAGGCTGC CTCGAGTGTG GGGGATAGGG  
10051 GGTCTCTGA CCCTAAGAGG CTGACCTCTT CTGACTGGG AATGTTGAC  
10101 TTTATAGCCA CTGGGTCACT CTCAGGTCTT AGGCCACAG TCCAGCTTGC  
10151 ATGCCTGACT GCACTTGGTC CCCGTCCCCC CCAGCCCCAC ACTGGTTCT

## FIGURE 3C

10201 AATCCTGTCC CCTCCCTGCA GGAGATGATC GCTCTCGGCT GCAGCAACTT  
10251 CTTGGCTCC TTCTTTAAAAA TTCATGTCAT TTGCTGTGCG CTTTCTGTCA  
10301 CTCTGGCTGT GGATGGAGCT GGAGGAAAAT CCCAGGTGAG CCTTGTCTA  
10351 GGGGAGTTGG GGGGAGGTGG TAAGAGAACAA GTTGCCTTCAA AAAAGCCTGG  
10401 GCACTGCAAG CCAGGCCAGC TCTTCTCCGA CCCCTTCTTC CCGTACTTAG  
10451 TCTCCACTCC ACCAAAGCCA TGGATTGGAA ATAATCAAG AGCAAAATT  
10501 TCACACCTTC CCTCTATCCC CAACTCTTC TCGGAATAGG TGGCCAGCCT  
10551 GTGTGTGTCT CTGGTGTGTA TGATCACCAT GCTGGTCTG GGGATCTATC  
10601 TGTATCCTCT CCCTAAGGTA AGAGCCCAGC CATCGAGCAG AAGTCAACGA  
10651 AAGACTCCAA TAAGAACAAAT CCCTGAGAGT TGTGTGGCAC TTTACGGACC  
10701 ACAAAAGTGC ACTGTTGTCA TACTTAGTCT CAAACACAAA CTGTGAGGTA  
10751 GACAATGCAG GTTTTATCCT CCCCCATTAA CAGGTGAAGG AACTGAGTC  
10801 TGAGAGTCTA AGTAACCTTG TCCATAGTGA GGCAGCTTAC AGCGCAGGGC  
10851 TGGTCCAAA CTCCAGCCTT CTGGCCTAG AGTCTAATCC CTAGGCAACA  
10901 TTTGCACCTA CCCACGAGTA CCAGGCTTT ATATAGCCCA GCTAGGAGGG  
10951 CTCTAGGCAT GCGTCATTTA GAGATGAGGG AAGAGAGATA GGGAAAGGAT  
11001 GGGGCCAGGA AGGACCCCAT GGCTCTAACG CCAGCACTTT CCAAACCTAA  
11051 GGTGAATGC AGAGATTTGG AGGATCAGCC AGGGGAGGTG TTCCAGAACT  
11101 CCGTCTCTGT CCTGCCAGGC CTTGGGTCG GGTATGCGCA GGAGGGCAA  
11151 AAGAAGGGGA GACCCCTGGGG TCCTGGAGCA ATGTTCTGCT TCTCTAGTCT  
11201 GTGCTAGGAG CCTGTATCGC TGTCAATCTC AAGAACTCCC TCAAGCAACT  
11251 CACCGACCCC TACTACCTGT GGAGGAAGAG CAAGCTGGAC TGTGTAAGTA  
11301 TCGGGCAGCC TCTGGGACT GGCCATGCC CTGCCCTCTC CTCCAACCCC  
11351 ACAGCCCTGT CAGCCCTGTC CTAACAAATGA ACCCTCTAGT CTGCTGCTTC  
11401 CTAATTAGCA TGAGATGAGT GGTTAAAGT CCGAGTTTCG AAGTAAACA  
11451 TCCTATGTT AAACCTAAC TCAGCCATCT GCTGGCTCCA TGGCAATAG  
11501 CAAGCCCTT AACCTTCCC AGTCTTGGTG TCTTAACTGG GCAAATGGTT  
11551 ATTTTATGCT CTCTGCCCTCC CAGGGTTTC TATGAAGAAG AAGCAAGGTA  
11601 ATACAAGTAA ACATGTTGTC TACATCGTAT TTTTAACTCA ATAAAGCTTA  
11651 GCTATGACTA CTTTATGACA TACAGCTTA AAAAACAAAAA GGAAATAGTT  
11701 TGTATTAA AAAAAAACCT AGAACATAAA GCCAGAGGAC CAAATCTG  
11751 AGCAAGTTAC TAGACTTCCC TGGGGTTCTA TTTCTCTATC TGTAAATGGG  
11801 GGTGAGACTC ATGCACTGAT GGTTGCGTC AACGCTGGTT CCGAGGATTA  
11851 AATGAGATCC CAGTGGGAAA ACACCGCATG AGCGAAACA TTCTGCAAAC  
11901 ATGACTTATT GTCCTGATTA GTCAACACACT CCACCGCATC ATCCGCTGG  
11951 CATACTAATG AAGGCCAGTG TGTGGTACG ACTGTGCCCTT CTCTCCATT  
12001 AAGCCCCACC ATAACCTATG GGAGAGGATT TACTAAACTT TCTTAACGGT  
12051 GATGAAACCA AGGCTCAGAA TGGTTAAAGTA AATTGTCAA GGCCACAGAG  
12101 GTAGGGAGTG GTAGAGTCTG GATTTAAACT CCAAGTCTG GACTCCAGAC  
12151 CTCTAGGCTG TACTGTCTCA TAGGGAAGGC AGTCTCACCC ACCTAGGGCA  
12201 GAGAAGAAAA TCCTTAAAGC CAGAGAAGTG AGTGGCTCAT CTGTGGTCAC  
12251 CCAGAGAGAC AGTGTATGAGG ACAGGGAGAA AAATTATACC TCAGTTCCA  
12301 GCCCAAGGAT CTGCTTGTAC CATAACCCAA CAAGCCCCCG CTATGGTGGT  
12351 ATTTCTCTAG GTTCATATGG CGGCTTTGT TTCCATTGTA TCTTCACAGC  
12401 AATTCTCTAC AGGAATCTGG CGACATTAT TTCCCTTGTAGA GGAATTCTCA  
12451 GGTCTAAAAA TCTATAGGGG GCAACTATCA AAACCTTCAAC CAATGTTGCC  
12501 CCCTACCCAC ACACAAAACC AGGCCCCAG CCGATCAGAA AGCACTGCTG  
12551 AGCTCCTGTC AGGGCCCAGC CAGCTCGCTG TGAGACAGAG AGAGGAAACT  
12601 CACATTATT GATCACCTAC TGAGCATCCA TCACTAGGCT AGGACCGTCA  
12651 CATTCTCTAA CTTTGTATC CTTTCATGAG GTAGGCATTA TTATTCTCT  
12701 TTTGTTTAC ATAGCCATTAA AAGAACAAAA TTGGGGCTG GGTGTGCTGA  
12751 CTCACACTG TGATCTAGCA CTTTAGGGGG CTGAGGCGAG AGGATCGCTT  
12801 GAAGTCAGGA TTTCAAGGTC AGCTGGGCA GCTTAGCGAG AGCCGCTCT  
12851 AGAAAAATAT AAAAGTTAGC TGGGTGTGGT GGCACGTGCC TATAGTCTA  
12901 ACTATTCAAGG AAGGTTAGGC GGGAGCACAA TTGGGGTTCC AGGGTTTGAG  
12951 GCTCCAGTGA GCTGATCTTG CCACTGCACT ACAGCCTGAG CAACAGAGCA  
13001 AGACCCCTGT ACTCCAAAAAA CAAACAAACAA AACACATTAA GAACCCAAAC  
13051 AGATCTGACC CAAGATGCA GCTCTTATAG ATGCCACCTC CCTGTGTGCT  
13101 GGGGCTCTA CTAACAAACAC AGACAAGATC AGGCAACCCAG AGTCAATCTA  
13151 AGGGAAAGAG GAAAGTGTAA CCAAAGCACA AATACATAAA ATATTGCAA  
13201 AATGCTATT AAAGAAAAAA AAGAGAAGAG AGGCTCTGAG GTTGTACTAA  
13251 CAGAGAATGG CTTGGCTAA TCCAGGAAGA CTTCCTGAAA GAGGTTGTTT  
13301 TTTCCCCAGG TCTGCTTTTG ACATCTCTCT TTTCACAGTG CATCTGGTA  
13351 GTGAGCTTCC TCTCCTCCCT CTTCCCTCAGC CTGCCCTATG GTGTGGCAGT  
13401 GGGTGTGCCC TTCTCCCTGC TGGTCGTGGT CTGGCAGACT CAGTTGTAAG  
13451 TGATAGCTTC CGCCCTCTTA GGCCCCACAGT CGGTTCCCTG GGCCAGCCCG  
13501 CAAAGGGCTT CCATGCCAGC GCCTGGCTTA GTCCACTGTA CCTTCCACCT  
13551 CTGGGCCTGG CACTGGAGGT GCTGCCAGGC CCAAAGAGAG CCCAACCCAG

## FIGURE 3D

13601 CCAGGACTGT GGGCACAGTC TGGGCTGTTT GACTTCCAT ATCTTGAAAA  
13651 CCCAGAGAA AGCCAGCATA CTCTTGTGG GGATGGCTGG GGAGAGGGCA  
13701 GTGGCAGAGA AAGGGGGCA AGGGCAGGTG GTGAGATTCA ACATCCTTCC  
13751 AAAGACATTG CCAGAACCCC AAACCAAATG GGACCCCAAC CCAGGAGAGC  
13801 GCCAGGGTGG AAGACAGAAAG CTGTGTTCTA CACACTGGGA GTATTACAGA  
13851 GAAGGGGTCT TGGCCAAGGC AGGGAGTACG CTGAATGTTG GGGGAATCCT  
13901 ATCTTCTCTT CTTGAGAACT CAGAACAAAGG AAATGATGAC TTCAGGGCA  
13951 CTCCACACAC TTCTCCACCC ACTTCTCTCC CCTGCCCTGT GGTCTGGGAG  
14001 CTATGTCAG GACCTGCCTG TCATCCTCAT AGTTATAGGA GGCCACAGGC  
14051 CACAGACAT GTGTCCTCC TGCAAAAAGA CAGACACAGC AAGTCTGGG  
14101 GTGAGGACAG GACCCCATCC TACCTTGGCT CTGCCCCCGC CCCAGCAGGG  
14151 GCACCCCTCC AGGCCCCATGT GCCATTAGCA TTCTCTTATG TTTTCTCTT  
14201 CCTGCTTCAT CCAGTCGAAA TGGCTATGCA CTGGCCCAAGG TCATGGACAC  
14251 TGACATTTAT GTGAATCCCA AGACCTATAA TAGGGTAGGT AATTCAAGCT  
14301 TATGACCTCC TTCTTTTGT CTGACCACCC CCAAGAAGAG GTTGCTTTT  
14351 AAAGCCAATA AAGACATTT TGCAACTTGA GCTCAGTCTC CCTGTCACAG  
14401 GCCCAGGATA TCCAGGGGAT TAAATCATC ACGACTGTGCT CCCCTCTCTA  
14451 CTTGCCCAAG TCAGAGATCT TCAGGCAAAA GGTCTACGCC AAGGTAAGGC  
14501 TCAGTCCCTG GCGACCAAGAG GCTCTGGACA GAGAGTGGCC GGAAAATGGA  
14551 AGCAGAAGGG CGGTGGGAGC TGAGAATAGG CCACTCCAT AGAGGGTGG  
14601 GGTCAAGATT GCTGTTGGCT CTCTCCCTGC AGACAGGCAT GGACCCCCAG  
14651 AAAGTATTAC TAGCCAAGCA AAAATACCTC AAGAAGCAGG AGAAGCCGAG  
14701 AATGAGGCC ACACAAACAGA GGAGGTCTCT ATTCTATGAAA ACCAAGGTGA  
14751 ATGAAGGCCA AAAGCAGCCC CGTGCCTCTG TCTCTGCCC ATTCTGATAC  
14801 TGCCCCCTGT TACTCATGGT ACCCTGGGGG CCCCCGCTTCC CACCTGACA  
14851 GGCAAGACA GAAAGTCTCT GGGAACACTG CCTGGTGGCC GCTGGCATT  
14901 TTTCTTCTTT TTTTCTTTT TCTTTTCTAG GATGGAATTT TGCTCTGTC  
14951 ACCCAGGCTT GAGTGAATG GCGTTATCTT GGCTCACTGC AACCTCCACC  
15001 TCTGGGGTTC AAGCGATTCT CCTGCCTTAG CCTCCCAAGT CGCTGAGATT  
15051 ACAGGTGCCA CCACACCCAG CTAATTTTG TATTTTTAGT AGATATTGGG  
15101 TTTCACCATG TTGGCCAGGG TGTTGTCAAA CTCTGACCT CAGGTGATCC  
15151 ACCTACCTTA GCCTTCAAA GTGCTGGGAT TACAAGCCTG AGCCACTGCG  
15201 CCCAGCCTGG GCATTTTCT TCTTGATGA GGTGCTACCA TCTCCAGGG  
15251 AAGCCACTGA ACCCCCCAAGG CCCTTCTCCA TTTTCTGGCT AAGATAGGAC  
15301 ATGGGCCATG GACTTTGAA CAACCCAGAG GGGGAACAGC AGTGAATTTC  
15351 CTGGGAACCC CAGGCAGCCC AGGGCTAGCA AGGCTGGGGT GGCCATGGCA  
15401 GTAATCCTTG TAATCCCGAG ACTTTAGGAG GCGCAGATGG GAGAATCACT  
15451 CTATGAGTT CAGGAGTTCG AGACCAGCT GCCAACAGTG GCGAAACGCT  
15501 GTCTCTACTA AAAATACACA AAAATTAGCC AGGCCTGGTG GTGGGCACCT  
15551 GTAATCCCAG CTACTCAGGA GGCTGAGGCA CGAGAACATCAC TTGAACCCGG  
15601 GAGGCAGAGG TTGCACTGAG CCGAGATAGT GCCACTGCAC TCCAGCCTAG  
15651 GCAACAGAGG GAGACTCTGT CTCAAGAAAT AAAGGAGCTC AGTGTCCCCG  
15701 GAGGGGCTTT CTCCCAGAGA GAGTGGGCTT GAGGCTTCAG TGCTCTCTT  
15751 GGCTGGGTCC TCTGACTTTG TCTGGGTTGT AGGAGACCAA GTTGCAGGC  
15801 CCTGCTTAAG AAAGGGCTTT GGGAGAGGCC TCTCTGGTGG AGCTTCAGG  
15851 GTCTGTGTT ACCATCACCG AGGCAGTTA TTCCCCCTACA CCTACACCC  
15901 CCATGCCCT GCTTCAGTCA CAGCAAGTC TGCTCAGTC TGGTGGTCCC  
15951 TGAECTGCCC CACTGTCCCC ACCCTTCCAG ACTGTCCTCC TGCAAGAGCT  
16001 GCAGCAGGAC TTTGAGAAATG CGCCCCCCCAC CGACCCCAAC ACAACCCAGA  
16051 CCCCGCTAA CGGCACCCAGC GTGTCTTATA TCACCTTCAG CCCTGACAGC  
16101 TCCTCACCTG CCCAGAGTG GCCACCCAGC TCCGCTGAGG CCCCCGGCGA  
16151 GCCCACTGAC ATGTCGGCA CGTCCCACCC CTTCTGTCACC TTCCACACCC  
16201 TCATCCTGGA CATGAGTGGA GTCACTTCTG TGACTTGTAT GGGCATCAAG  
16251 GCCCTGGCCA AGGTGAGGCC CTCGGGACA GCAAGCACCA CCCACTCCAC  
16301 CCCCTCCGCT CTGCTCTCCA CATTCCCTT CCTGGGAGCC CTCATTTCTAG  
16351 GAAGCTGAGG GAGGAAGCTC ACTGGGAGA CTAAACAGCTC CTAGGAATCC  
16401 CTCTTTCTCC CAGACGCCAC CAGGGTGGAG CATTCTCCAC AGAGCAGGCC  
16451 CAGACGGGCC ATGACAATGA GTGGGGGGAC AAGTCTACCA GAGTTTCAGG  
16501 CCCCTGTGCT CCCAACACCC CCAGCAGTGG CCATCCCAAG TCCCTCTAG  
16551 CCATCAGGAA CCCACCCAGG TTCTCTGAGG AGGGTCCAGT TTGGCTCTG  
16601 GTTCATGATC TGCTGCCCTT GTCCCTCATT CACCAGCCAC CCTAGGACAG  
16651 GAGAAGAAAT AATACCACTG CCCCCACACCA TCAGGCCAAA CAGAGAGCCC  
16701 ACGGGACACC TTGAATGAAT GTATCCATCT GATAACTTT CAGCAGGCCAC  
16751 CGCCAATGGC GGGAGTCAGC AAACCTCAGA GCTGGCTCAG ATAGAGGCAA  
16801 GCCAGGGGAA CAATGGGCAC AGAGAGTGT CGGACTGCCT TCACCATCAA  
16851 CCAGGCAGG GGCAGGGCCC ATACCCAGCC TTGGGCCCTCA GCCGGCTTCC  
16901 TTAGCCAGGA TCTGGAGTCC AGGCCAGCT TGCTGAAGC TCTAGACTCC  
16951 CTGAGCCTCC ATCCTCCCT GCAGCTTCTG TCTGAAGCCA CAAAGAAGTC

FIGURE 3E

17001 TGAGAATCTA AGCTACTGAA AGAAAAGATC AGCCGGGCGT GGTGGCTCAC  
17051 TCCTGTAATC CCAGCACTTT GGGAGGCCAA GGCAGGTGGA TCACAAGGTC  
17101 AGGAGTTCAA GACCAGCTG GCCAACATGG TGAACCCCG CCTCTACTAA  
17151 AAATACAAAA ATTAGCCAGG TGTGGTGACG GGGCCCTGTA GTCCCAGCTA  
17201 CTCGGTAGGC TGAGGGCAGAG AATTGCTGAA ACCCAGGAGG CGGAGGTTGC  
17251 AGTGAGCCAA GATCGCGCCA CTGCACTCCA GCCTGGGCAA CAGAGTGAAGA  
17301 CTCCATCTCA AAAGAAAAAA AAAGAAAATA TCTAGCCCCA CAAGAAGGGG  
17351 CCATGGTGAC TTTAAGTGCC CGCCACGTT GCAAAAGTCC ATTTCCGCTC  
17401 CACTTCCCAG AGAAACCGTC AGCCAACACT CCAGGGAGAA GTGGTGTGCT  
17451 TTGCTGCTAT TTTTGTCTTT GGCTGCTGGG CTCTCAGGGT TGCTTATTG  
17501 TTGGCTTCC CCTCTGAAGT ACGTTTGTG AATCACTTT GAGACCCACT  
17551 CAGAACATTC CTTTCCCTTT GCCTCCCTAC CCCAACAACA CTTCTAGCTG  
17601 AGCTCCACCT ATGGGAAGAT CGGCGTGAAG GTCTTCTTGG TGAACATCCA  
17651 TGGTAAGAGA AAGAGGACAT TTAGGGACTG AAAGACTGGC AAGGAGTGTG  
17701 GGGTAGGAAC AGGTGGTGG GGCTGAATA GTGAGGAGGT TGGAAACGAG  
17751 AGCACCCAGC TATCCCCCA AAGCTGCTGC CTGCTCATAA AAGCTTCAGG  
17801 TACAAGTCCA AAGAGACTGG TCAGATTGCA TAAACATCCT AGGGGCCTTA  
17851 GTGACAGAGT GGGGGTGAGG AGGTCACTGAA GTTACAGAAG GACAGCTAGG  
17901 ATTCTAATCT ACCCCATAAC TAATTGCCA CGTATCCTT GCGGAGTCAC  
17951 TTTATCTCTC AAGGGATCTA TTTCTACCTA TGAAAACGA GAGGGTTGAC  
18001 TAGATGGATT TGGGGATCCT CTCCCAATCA GAAACTCTGT GAATCGATAT  
18051 AGGCATAGAG CACACGGTAC CCTAATTCCC CAGGGAAACAT ATAAATATGC  
18101 AGTTTGTGAT GCATACAGCC TCCAAAGGGT GCATATACAC AGCCTCAAGG  
18151 ACCTGGCCAC AGGGCAGCA ACATTTACAT GACTAGCATG TACGCAAAGT  
18201 GCAGAGATGT GGGAGCAAGT GCACACAGAC ACACAGGAGA ATGTGAAGGG  
18251 GCACATACAC ACACACCCAG CTCCCTGCAC TGGGTAGAC CCCCTCCAGC  
18301 AGGGCTGCAG TTCCCAAGCT CCGCATGGCC ACGTTGGGG AGAGAATCTG  
18351 CAGTGGCAAT GACCTGCTAT GATATGTTCT GGAGTTAGAA GCAGTGGATT  
18401 CTCCCCAAC TCACTGGACA CCCCTTAGG AAACCATCTC TAGGATTAAG  
18451 AGTAATCCAC ACAAACTTC AATGCCACAC ATTGGAAGTT GCTGGAAGG  
18501 TCTGGAAAAA CAAGAGGAAG GATGGGCTCT TGGGGATAG AACTGGCAGC  
18551 GGCCTCTCA AGGATGGCTT AGGCTTTCC ACTCGAATCA CCACAAAGTA  
18601 CTGACTCCCT AAATCAAAC TCTTCTTCT GCTCTGGTT GAAACTTCAG  
18651 CATCCTCAAG TTTCATGTTGC CCTCTGCCGT CCAGAACTGA TATTGCACTG  
18701 CCAATGCCAT GGGCCCTCAGA TACAGCAAGA GCTGGGACCT CAGGCCTCTC  
18751 CCATCCCTGC TCTGGTCTCA CTATCTCCC CACCCCCAGC TCCAATCCAC  
18801 AATGGCTGTT ATCTTCTGA AGGTGATCTT TTCTCTTCT AGCCCCAGGTG  
18851 TACAATGACA TTAGCCATGG AGGCCTCTT GAGGATGGGA GTCTAGAATG  
18901 CAAGCACGTC TTTCCCAGCA TACATGACGC AGTCCTCTT GCCCAGGCAA  
18951 ATGCTAGAGA CGTGAACCCA GGACACAAC TCCAAGGGGT AAGGTTCTG  
19001 CACCTGGGGA ATCCTAGGCT CCAAGGCACT GAAATAGCAG GACCAAGGAG  
19051 CATTATTAGA AAGAACACAG GAGAAGGTTT AAGTTCCAAT ATCAAGTCTG  
19101 CCATTTCAGT TTTCATGAAATC TGTTTCTTA TCTATAGAAT GAGCACCAC  
19151 AACTAACATT ACCTACCTCT CTGCATTTT CTTTATTTT GTTTTAGGGT  
19201 TAAATGATAA TTACATCTTT TGTGTCACTT GAAAGCACTT TGTGTATTG  
19251 AAAAATTCTT TATCAATATA AGTTTCTGG TTGCACAAAC ACCCAAAGCA  
19301 TAGTAGAGCA GGCCCACCTCT GCTGGCATCG TCCCCTGCCT CCTCCTCATC  
19351 TCTTTCTAAA GGGGGCTTTC GGGGAAGGGAG GGGAGGGGAG TAAGCCTACC  
19401 CATTTTAACT TACCGGAGCT TAGAGATTC AGGCTGGTGA GGGATAAAGA  
19451 GATGGGGTCT GAGTTTTGTC TCAGCTTTT GACATTTAAT TTACTAGCTC  
19501 AGTAAGTCAT ACAAAATGGGA TACAAATAAC ACCATCTAAA ACTCCAGAAG  
19551 ACTGGGGAGT CAGAAAAATC CTACCTCTT GGGGTCCCTG CCCAGATCCC  
19601 CAGTCATCTC TAGCCCTCA GGTCCCCCTCC CAGCTCAGCT CCTGCCCTTG  
19651 GCCTCCAAG ACTCTTGTG TGCCCCAGCC CTGGGTAAAA ACCTCCCTG  
19701 CCCTCTGTGG GTCTATAAGAA AGGCTTTCT GGGCCTAGAG CAATGATTG  
19751 CTCTTTGCCT TAAGAGACTG ATGAAGGTTGA AACCATCTGT TCTAAGTGT  
19801 GAAAGACTGC CCAGGAACAC ACAGGGCGCT GGCTCCTGCC CTCCATGCCT  
19851 AGAGGGAAAC CCTGGGGAAA CAACGGGCTT TCCTGCTTCG TGAAATTG  
19901 CCGCAGAGCA AAGAGGGAGA TTCTGGAGGA AGTCGATTA GTTGTAGTG  
19951 CCCTTAATCAT GTTCAGCTAC TCTAGTTGG ATGTACTT GATTAGTCAT  
20001 AGCACTTATA ATAATTTTAT ATTATATACTT ATATATACTT ACATATTATA  
20051 GACCATTAC AGATACAAAT CACACACATA AACACACACC TTTTCAACAG  
20101 CATTGTGAGG GACAAAGCA GCAAAGTGGAG GCTGGTTATC AGACTTTAAC  
20151 AGATTAGAAA ATATATTCCC AGGAGGACAG GAATTCACCA AGGTCAAGGCA  
20201 GCTAGCCAAT AGTTTTCTA AGCTGAGTAA AACCTTCCCT GCCTCTAACG  
20251 GCCCACAAAG GAGGGAAAGAC CGCGATACAC ACCTGTCTGG TATAAGGGGG  
20301 AAGACCACAG CCGTGTGTT TTTGTGAGGC AGGTAAGGGG AGGGGCAAGA  
20351 GGATAAGTCA TGTGTCAGGA AGCAGCGTCC AACAGAGGCC GGCCACCTGT

## FIGURE 3F

20401 CCCTTTCCT GCCACCATGC ACCAACTTTG CTGTTCAGTC ACTGAAGCTC  
20451 ATTCTGCACT GGCTTCTTC CTTCCAGGCT CCAGGGGATG CTGAGCTCTC  
20501 CTTGTACGAC TCAGAGGAGG ACATTCGAG CTACTGGAC TTAGAGCAGG  
20551 TGAGCTGAGG GAAGGGGCTG TGAGGGTGGG AGCAGGGCGA AGAGGGGAAG  
20601 GATGGGGTCG CTGTCATAA CAAGGCAGTC ACTCAGCTGT CTCACCTCCA  
20651 GCCCAGAGCA GTCACATTCA AGGCCACAAA GATTGTGGT CATCTTGT  
20701 TTTTTCTTT TCCTTTCTT TTTTTTTTT TTTTAATTG AGACAAAGTC  
20751 TCACTCTATC ACCCAGACTG GAATGCAGTG GCATGATCTC AGCTCACTGC  
20801 AACCTCTGC TCCCAGGTT CAGAGGTTCT CTCAGCTCAG CCTCCCGAGT  
20851 AGCTGGACT TCAGGGCTGC GCCCAGCTAA TTTTTGTATT TTTAGTAGAG  
20901 ACAGCTTTTC ACCATGTTGG CTGGGCTGGT CTCGAACCTC CGATCTCAAG  
20951 CAATCTGCCT GCCTCGGTCT CCTAAGTGCC TGAGATTACAG GCATAAGCCA  
21001 CGATGCCTGG CTTTGTCTT CATTCTCTC ACTCCCTGAA AGGCATCGTG  
21051 GGGAGAGGGT GAGTCAGTGG ACCAAGTCCT AGAGAACAG TATCTATTCT  
21101 TATTCTCAA ACATCACCC ACGTGACCCCT GAGCAAGCCA CATACACCC  
21151 GGGCCCTAGT TTTTATCATC TGAAATTG GGGGAAACAT AGGTAATACC  
21201 TGTCCCATCC ACCACAAAG ATTGGCAGGG CAGTCACTTG TTCTTCTATT  
21251 AATTCAGCAG GTATTTATGG CGTACCTACT GTTTGCCTGA CACAGTTCA  
21301 GATGGGCACA TAGCAGTGG CAAAACAAAG GCCTCTGCCT TTTAGAAACT  
21351 TACGTTATGG TAGAATAGAT GGATTTNNNN NNNNNNNNNNN NNNNNNNNNNN  
21401 NNNNNNNNNNN NNNNNNNNNNN NNNNNNGTCT ACAAAATGAAT TATTATTGCA  
21451 TGTGACAAG CTTAAAGAAC TAAAAAAAT GTGGCTGGT GCAATGGTC  
21501 ACACCTGTAA TCCCAGCACT TTGGGAGGT GAGGTGGCG GACCACCTGA  
21551 GGTCAAGGAGT TTGAGACCA CGTGGCCAACT ATGGCGAAC CCCGTCTCTA  
21601 CTAAAAGCAC AAAAATTAGC CAGGCGTAGT GGTGCATGCC TGAGTC  
21651 GCTACTCGGA AGTCTGAGGC ATGAGAAATCA CTGAAACCTG GGAGGCAGAT  
21701 GTTGAGCTGA GCCGAGATCG TGCCACTGCA CTCCAGCTTG GGTGACAGAG  
21751 CTAGACTGTC TCAAAACAA ACAAAACAA ACAAAACCTAA AAGATATGTG  
21801 GATATGAGGG ATCACCATCC CCATAGGGCC CCTGGATTAA CACCAACCCA  
21851 CCAATGCCCT GAATTTAAAG AAACCGATG ACTAGGTTTG GAGAAATCTG  
21901 GCTTTGGTC TATGAGAAGT AGTGTCTCTC TTGTCCTGC TTCCCATTCT  
21951 TTTTGACATT GAGCTCATG TGCTCTGAA TCCGTCCTCTC ACAGTGTGA  
22001 TGGCAGGTGG GACAGATTAG AAAATAGAGC TGAGGCCACA GAGATTTGGC  
22051 AGACTGATT CGGTGCCCT TTGGAATCTC CAGCACATT CAAAAAGCCT  
22101 GGATAGGACC AAAAATAGCTT ATCAACGTGA GAAAGGACTT CAGAGCTTGT  
22151 CTACTGCCAA CCCTCATTTT ACCCAATGAG GAAAGTGAAG CTATTAGGG  
22201 GCGAGGGACA CGTGGAGGT CACACAGCAC ACAGGAGGTG ATTCACTGT  
22251 AGATTTAGC ACCTGCTCT GCCACGCTGG ACTGGTTCAC CTCTTAGGCT  
22301 GACCTGCCT CTCCCCCTGTT CACACACACT CTCGCACACA CACACACACA  
22351 CACACACACA CACAGGTGCT TTGTTCTGGC CAGGGGTTCC TAGGGTCACC  
22401 TCTTGGTGC AGCCACTGTG ACCCCAACTG GTCTAACCTC TCTCTTCCCC  
22451 TCCCACCTTC TTCTCTGTGTT CTCGAGGA GATGTTGGG AGCATGTTTC  
22501 ACGCAGAGAC CCTGACCGCC CTGTGAGGC TCAGCCAGTC CTCTAGCTGC  
22551 CTACAGAGTG CCTGGCACTT GGGACTTCA TAAAGGATGA GCCTGGGTC  
22601 ACAGGGGTG TCGGGCGGAG GAAAGTGCAT CCCCCAGAGC TTGGGTTCC  
22651 CTCTCTCTC CCCCTCTCTC CTCCCTTCTC TCCCTCCCCG CATCTCCAGA  
22701 GAGAGCCTCT CAGCAGCAGG GGGGTGCTAC CCTTACAGGA GTGAGAGTCT  
22751 GGTGAGCCCA CTCTTACCC GTCAGGCCCT GGGCGCAATG GACAAGCCTC  
22801 CTGCTCACT CACCCCCACCC ACCTCTGCC TGTCTTGGC AGCTGAAGGA  
22851 CACCTTGACT TCCAGCTTTT ACGAGTGGC CAAAAACAGA AGGACAAGTA  
22901 CAACTGTGCT GGCCTGCTGT ACAAGCTTC AAAAGTGTCC CAGAGCCAC  
22951 ACGGCTCGGT GTCAAGATGGT GTCAGGCTGT CACGGACATA GGGATAAACT  
23001 TGGTAGGAC TCTGGCTTGC CTTCCCCAGC TGCTCAACT CTGTCTCTGG  
23051 CAGCTCTGCA CCCAGGGAC ATGTGCTCTC CACACCCAGG AGTCTAGGCC  
23101 TTGGTAACTA TGCGCCCCCCC GTCCATCATC CCAAGGCTG CCAAAACAC  
23151 CACTGCTGTC AGCAAGCACA TCAGACTCTA GCCTGGACAG TGGCCAGGAC  
23201 CGTCAGAGACC ACCAGAGCTA CCTCCCCGGG GACAGCCAC TAAGGTTCTG  
23251 CCTCAGCCTC CTGAAACATC ACTGCCCTCA GAGGCTGCTC CTTCCCCCTG  
23301 GAGGCTGGCT AGAAACCCCA AAGAGGGGAG TGGGTAGCTG GCAGAACAT  
23351 CTGGCATCTC AGTAATAGAT ACCAGTTATT CTGCACAAAAA CTTTGGGAA  
23401 TCCCTTTG CACCCAGAGA CTCAGAGGGG AAGAGGGTGC TAGTACCAAC  
23451 ACAGGGAAAA CGGATGGGAC CTGGGGCCAG ACAGTCCCCC TTGACCCAG  
23501 GGCCCATCAG GGAAATGCCT CCCTTGGTA AATCTGCCTT ATCCCTCTT  
23551 ACCTGGAAA GAGCAATCA TGTTAACTCT TCCTTATCAG CCTGTGGCCC  
23601 AGAGACACAA TGGGGTCCTT CTGAGGCAA AGGTGGAAGT CCTCCAGGG  
23651 TCCGCTACAT CCCCTAACTG CATGCAGATG TGAAAGGGG CTGATCCAGA  
23701 TTGGGTCTTC CTGACAGGA AGACTCTTA ACACCCCTAG GACCTCAGGC  
23751 CATCTCTCC TATGAAGATG AAAATAGGGG TTAAGTTTC CATATGTACA

## FIGURE 3G

23801 AGGAGGTATT GAGAGGAACC CTACTGTTGA CTTGAAAATA AATAGGTTCC  
 23851 ATGTGTAAGT GTTTGTAAA ATTCAGTGG AAATGCACAG AAAATCTTCT  
 23901 GGCCTCTCAT CACTGCTTTT CTCAAGCTTC TTCAGCTTAA CAACCCCTTC  
 23951 CCTAACAGGT TGGGCTGGCC CAGCCTAGGA AACACATCCCC ATTTCTAAT  
 24001 TCAGGCCAGAC CTGCGTTGTG TGCTGTGTG TTGAGTGAGC TGGTCAGCTA  
 24051 ACAAGTCTTC TTAGAGTTAA AGGAGGGGGT GCTGGCCAAG AGCCAACACA  
 24101 TTCTTGGCCC AGGAGCATTG CTTTTCTGTG AATTCAATTAT GCCATCTGGC  
 24151 TGCCAATGGA ACTCAAAACT TGGAAAGGCAGA AGGACAATGT TATCTGGGAT  
 24201 TCACCGTGCA CAGCACCCGA AGTGCCTAAAT TCCAGGAGGA CAAGAGCCTT  
 24251 AGCCAATGAC AACTCACTCT CCCCTACTCC ACCTCCTTCC AAGTCCAGCT  
 24301 CAGGCCAGG AGGTGGGAGA AGGTCAACAGA GCCTCAGGAA TTTCCAAGTC  
 24351 AGAGTCCCCCT TTGAACCAAG TATCTAGATC CCCTGAGGAC TTGATGAAGT  
 24401 GATCCTTAAC CCCCCAAGTAA TCATTAACCC CCAGACCAGC CTCAGAACTG  
 24451 AAGGAGATTG TTGACCCAGT GACCTGGAGT TGAGGCTCAG GGAGAGATCT  
 24501 GCCACATGTC TGAGGGTTGC AGAGCC (SEQ ID NO:3)

**FEATURES:**

Start: 1997  
 Exon: 1997-2121  
 Intron: 2122-4732  
 Exon: 4733-4872  
 Intron: 4873-5004  
 Exon: 5005-5115  
 Intron: 5116-5781  
 Exon: 5782-5957  
 Intron: 5958-7770  
 Exon: 7771-7935  
 Intron: 7936-8470  
 Exon: 8471-8623  
 Intron: 8624-8917  
 Exon: 8918-9000  
 Intron: 9001-9777  
 Exon: 9778-9925  
 Intron: 9926-10221  
 Exon: 10222-10335  
 Intron: 10336-10539  
 Exon: 10540-10617  
 Intron: 10618-11197  
 Exon: 11198-11293  
 Intron: 11294-13338  
 Exon: 13339-13445  
 Intron: 13446-14214  
 Exon: 14215-14284  
 Intron: 14285-14400  
 Exon: 14401-14493  
 Intron: 14494-15980  
 Exon: 15981-16262  
 Intron: 16263-17597  
 Exon: 17598-17652  
 Intron: 17653-18842  
 Exon: 18843-18988  
 Intron: 18989-20477  
 Exon: 20478-20549  
 Intron: 20550-22478  
 Exon: 22479-22523  
 Stop: 22524

**CHROMOSOME MAP POSITION:**  
 Chromosome 1

**ALLELIC VARIANTS (SNPs):**

| DNA Position | Major | Minor | Domain         | Protein Position | Major | Minor |
|--------------|-------|-------|----------------|------------------|-------|-------|
| 48           | C     | G     | Beyond ORF(5') |                  |       |       |
| 132          | G     | A     | Beyond ORF(5') |                  |       |       |
| 724          | A     | C     | Beyond ORF(5') |                  |       |       |
| 1558         | C     | G     | Beyond ORF(5') |                  |       |       |

**FIGURE 3H**

|       |   |     |                |     |   |   |
|-------|---|-----|----------------|-----|---|---|
| 1577  | A | G   | Beyond ORF(5') |     |   |   |
| 2487  | C | A   | Intron         |     |   |   |
| 2634  | T | C   | Intron         |     |   |   |
| 4352  | A | G   | Intron         |     |   |   |
| 5157  | A | C   | Intron         |     |   |   |
| 5658  | A | T   | Intron         |     |   |   |
| 5945  | T | C   | Exon           | 180 | T | T |
| 6281  | C | T   | Intron         |     |   |   |
| 6452  | G | C   | Intron         |     |   |   |
| 6610  | T | G   | Intron         |     |   |   |
| 7247  | T | C   | Intron         |     |   |   |
| 7360  | A | G   | Intron         |     |   |   |
| 7644  | A | T   | Intron         |     |   |   |
| 8127  | A | C   | Intron         |     |   |   |
| 8317  | G | A   | Intron         |     |   |   |
| 9079  | G | A   | Intron         |     |   |   |
| 9537  | G | T   | Intron         |     |   |   |
| 12302 | C | G   | Intron         |     |   |   |
| 12354 | C | T   | Intron         |     |   |   |
| 12487 | C | T   | Intron         |     |   |   |
| 13198 | - | A   | Intron         |     |   |   |
| 13257 | A | G   | Intron         |     |   |   |
| 14541 | G | A   | Intron         |     |   |   |
| 14545 | A | G   | Intron         |     |   |   |
| 15041 | C | A   | Intron         |     |   |   |
| 15053 | A | C   | Intron         |     |   |   |
| 15065 | A | G   | Intron         |     |   |   |
| 15108 | A | C   | Intron         |     |   |   |
| 16274 | - | G   | Intron         |     |   |   |
| 17424 | C | T   | Intron         |     |   |   |
| 17627 | G | A   | Exon           | 657 | V | V |
| 18427 | T | C   | Intron         |     |   |   |
| 18813 | C | G   | Intron         |     |   |   |
| 19035 | T | C   | Intron         |     |   |   |
| 19182 | T | C   | Intron         |     |   |   |
| 19508 | - | G C | Intron         |     |   |   |
| 19571 | T | G C | Intron         |     |   |   |
| 20147 | T | G   | Intron         |     |   |   |
| 20180 | G | A   | Intron         |     |   |   |
| 20584 | A | T   | Intron         |     |   |   |
| 20717 | T | C   | Intron         |     |   |   |
| 20894 | A | G   | Intron         |     |   |   |
| 21787 | - | A C | Intron         |     |   |   |
| 22264 | T | C   | Intron         |     |   |   |
| 22338 | - | C A | Intron         |     |   |   |
| 23363 | T | C   | Beyond ORF(3') |     |   |   |
| 23688 | G | A   | Beyond ORF(3') |     |   |   |
| 24210 | A | C   | Beyond ORF(3') |     |   |   |

Context:

DNA  
Position

48 CTGGGTTCTATGTGGGGAGGTATGCTCCCCACTCATTGAGCCCC  
 [C, G]  
 CAGGCAAACCACCTGGACAGCCAGACCCATGCAGACTCTGGAGCAGGTGGAGAGGAAGAG  
 TGAGACCACCCCGCCTACGGGGGTGAAGGGCCGGCAGCCTCTGAATAGTCTCTGCTAG  
 GAGGTAGAAAGCACCCCTCCATCTTAATCATAGTAATCATGCCACTACCATTACTGGG  
 TGCCCTATAAAAGGCCAGCCTCTTCATACACATGATCTCACTGAATCCTCATAGCATCTGC  
 CTGCGACTGTTATTATCCCCATTACAGATGAAGAAACTGAATCTTGAACCCAGGTAT

132 CTGGGTTCTATGTGGGGAGGTATGCTCCCCACTCATTGAGCCCCCCCCAGGCAAACAC  
 CTGGACAGCCAGACCCATGCAGACTCTGGAGCAGGTGGAGAGGAAGAGTGAACCC  
 GCCTCACGGGC  
 [G, A]  
 GTGAAGGGCCGGCAGCCTCTGAATAGTCTCTGCTAGGAGGTAGAAAGCACCCCTCCATCT  
 TAATCATAGTAATCATGCCACTACCATTACTGGGTGCTATAAAAGGCCAGCCTCTTC  
 ATACACATGATCTCACTGAATCCTCATAGCATCTGCCCTGGACTGTTATTATCCCCATT

## FIGURE 3I

ACAGATGAAGAAACTGAATCTTGAACCCAGGTATCTGGCTCTCAAACCTGTGCTGTT  
 TCCCTAAGCCACCCGGTCTCTCATTCTCCACTGAAATGTCACATGCCATTGCCCTT

724 ATTGCCCTTACTCATTCTGCCCATGTCCTCTCCAAAACACCAATTATCAATTGCTCAA  
 CAAGTATGTGTTGAGTACACACTAAGGGCCAGGCAGGGGCTGGGCACAGGCCTGGGG  
 TAGGTTATTCTCCACCTCGTTCTGCTGGGTATCACCTGTGGGTCTTGCCGGCAT  
 CCCACCCCTACCTGTAGTTCAAGTGGACCTGGGATCCAAGACCAAATGAATGGAATGC  
 ACCAGGCCAGCCTTACCAACTTGAGCACAATCTTATTCTATAATAGAAAATCACATTG  
 [A, C]  
 TCACACTTACATTTCACACAACCCCTTATCCATTAACTCATTGATCTTACAACA  
 ACCCTGTGAGATATGTCGTTACTCCACTTTAGTGTACAGAATCTGAGGTTGAAAAG  
 TAATGTCGACCAATTGCTCTCATTAAATAAAAGCAGGATTAACCCAGGCTCTGGACCC  
 CCACAAAAGGCATTAAGCAACCTGCTCCCTCTGACAACCTCCCTGTCACCCAGGCTCT  
 CCTCTGGGAAGTTGGGGCATCTAGCCCCAAGTAGTTACTCATTTCACACCCATCT

1558 TCAGCTCTGCCCATCTCAGCTCTGGAACGTCAGCCAGGGTGCACAAAAGTGAGGAGGA  
 GAGGAGCGGCAGTACACAAGGGTGGGGAAAGATTAGGCACAGGAAGCCGTGGGAGAGAG  
 AGCCGGCAGGTGGACCATCTGGTTTCCCCACACACACATTGTCCTGGGAAACCTG  
 TTGGTGAAGTTCTAGATGTCCTATCCAAGAAGGGTCTTGTAGGTATCTCAGCTATCC  
 CCCTGCCTCTAGGCAAGCTGTTTCTGTTCTTCAAGCTGACTGGCTGAATGGTAGGAG  
 [C, G]  
 CTTCTGCAGGGAAACTAAGGCTGGGAAGGGAGTATGGCTTGTGGGGACACCAAGGGT  
 CAGGGGAGGGGAGGGTCCACCTGTAAGTGGGGCTCTGGCTGTGATTCCCC  
 TTGCTGGTCTAGTGGGGTGTGGTGAACGCCACAGGTGTGGAGTGCCAGCACGTG  
 CTGAGCGCAGCAAAACAGCCAGGGTAGTCTATGCATCATCAGTGCCTGGGAAGGAAG  
 GCCACTCGGAGCAGGGAGTCTGACGGAAAAACTTGACAGAGGAAGGGAGGCACCTTGCT

1577 CTCCCTGGAACGT CAGCCAGGTTGCACAAAAGTGAGGAGGAGAGGAGCGCAGTACACAA  
 GGGTGGGGAAAGATTAGGCACAGGAAGCCGTGGGAGAGAGAGCCGGCAGGTGGACCATC  
 CTGGTTTCCCCACACACACATTGTCCTGGGAAACCTGTTGGTGAAGTTCTAGATGT  
 CTTATCCAAGAAGGGTCTCTTGAGGTATCTCAGTATCCCCCTGCTCTAGGCAAGCT  
 GTTTCTGTTCTTCAAGCTGACTGGCTGAATGGTAGGAGCCTTCTGCCAGGGAAACT  
 [A, G]  
 AGGTCTGGGAAGGGAGTATGGCTTGTGGGGACACCAGGGTCAAGGGAGGGAGGGTCCA  
 CCTGCTGAATCAAGTGGGGCTCTGCCCCCTGTGATTCCCCCTTGCTGGTGTCTAGTGG  
 GGGTGTGGTACGCCACAGGTGTGGAGTGCAGCCACGTGCTGAGGCCAAGCAAAACA  
 GCCAGGGTAGTCTATGCATCATCAGTGCCTGGGAAGGAAGGCCACTGGAGCAGGGAGT  
 CTGACGGAAAAACTTGACAGAGGAAGGGAGGCACCTGCTTATGGGGCGGGGAAGGC

2487 ACACGGCTCTGCACTGGTATCCCTAACATGGGTTAACGGGAGGCCCTGGGAAGTGAGG  
 TTCTGAATGATGAATTAAAGATCTACAAACCTCATCTGACTGAGACCCCCAGGGAGGAT  
 GGGGAGCAGGAGCAAGAACCATCCAGAACGGTTATATGCCATTCCAAACCCCTGCATGG  
 CATCTCCCATATTCTCAATTACCCGGTCTCTGGGTTTGTAAAGCATGGTAGATGA  
 GCATCTACGTTATGGAGGGTGGGAGCATCAGAGCCCTACTCCATGCCCTGTTCCCTC  
 [C, A]  
 TTACAAAAAAACTCTGAAGTTACCATCACCCAGGTTCTTGTCTTCCCTCCGGATG  
 TTCTTCCCTCACTGGTCCAGAGAATGCCAAAGGGGCCCTAAATTCTGAACCTTCC  
 TGAGGGACCTACCAGGGTAGTGTCTACCAGCGCCCAAGGGTCTTCCACTCTCATCTC  
 CTGGAAATGCATGGTGGGTATGAAACCTTGTCCCTAACGGCGTACACAAGGTGATC  
 CATACCCACACCCAGGAGGCTGGGCTGCGGGTGTACCCCTCCCCATTCCAGACTCCT

2634 AGGGTTATATGCCATTCCAAACCCCTGCATGGCATCTCCATATTCTAACCCGG  
 GTCTCTGGGTTTGTAAAGGCATGGTAGATGAGCATCTACGTTATGGAGGGGGGGAG  
 CATCAGAGCCCTACTCCATGCCGTTCCTCTTACAAAAAAACCTGAAGTTACCAT  
 CACCCAGGTTCTTGTCTTCCCTCCCGATGTTCTCTCCACTTGGTCCAGAGAA  
 TGCCAAAAGGAGGCCCTAAATTCTGAACCTTCTGAGGGGACCTACCCAGGGTAGTCC  
 [T, C]  
 ACCAGCGCCCAAGGGTCTTCAACTCTCATCTCCCTGGAAATGCATGGTGGGTATGAAAC  
 CTTGTCCTAACGGCGTACACAAGGTGATCCATACCCACACCCAGGAGGCTGGGC  
 TGCGGGTGTCAACCCCTCCCACTCCAGACTCTGGCAGACCTCTGGCCAGCTATAG  
 GCCAACTCACTCTCCCTCACTCCCTGGGGAAACGGCTGATTCAAGTACCTGGATTGAGG  
 TCACTGGCAATGGCTGAAGTGGAGACGCAGGTGGAACTGGTTCAAGGCCGGGGAAATCACC

4352 ATTGGAGTTACACACATAAAGGATAGTGAGTCAGCAGAGTGACCCCTGCAGGAACAATA  
 GAGCCTCTTTCAAGGAAGTTCAAGAAAATGGCAGCAGGGCAGGGCCACTCGGGTG  
 TATTCACTATTCAACAAATATTACTAACGGAGAGCTAGACTTCTGCCATGCGTGGTGGGGTTGCTGCTA  
 GTGGGAGAGACAGACAAAAAGCAAGGAATGCACACACAGGATGCACACAGCGGCAGGA

## FIGURE 3J

[A, G]  
 CCAAGGTGCGATTACCGGCCCTGGGATCAGACAGACAGGACTCAGAGGAGACTTCTCA  
 GAGAAAAGCCATCTGAGCCAAGGGATGGATCTGATACCTCCGAAGGCTGAGCCACCTAA  
 CACTCATACCTTAAAGCAAGTCTTATAAACTCCCCAGGTAAAGCAGCTGCAGTCAGAAG  
 ACCTCCAGCTAATGCCAGGACAAGTTGAGCTCTCAAGAAAAGTTCTGCCTTTC  
 TTCTCAATATCCCTGGCACACAGTCAGTGAATTGAATGAACCAATGAATGAG  
 5157 ATCCAGGTCCCACAAGGTGAAGGGCTCCTTCAGCCAGGCCCTGGATTGCCACTCCCTCA  
 CCCTTCTCTCCTCATCCCACTCCATCCCTCTGTGATCCCCATAAGCTAGTCATGCTGC  
 TGAGCTTCAGTCTCGTGTCTCTGCAGGCATGGCATTGCTCTGCTGGCCAACCTTCT  
 GCAGTCATGGCCTCTACTCTCCCTCTCCCCCTCTGACCTACTTCTCTGGGGGT  
 GTTCAACAGATGGTGCAGGTAAAGGCCCTCTCCCCCTGGGCAGGCAGGATGACCCAGACC  
 [A, C]  
 CAAGGATGGGAGGTGGCAAAGGGGCCCTGGGAGATTTCATCTGCATTCTCTGGAG  
 TTGTTCTGGTCAGTCTAGGGAAATGGTCACTGTGAATGTCATTTCCAGGTCTCGGTG  
 ACCTTGGAGAAACCACTGAGCCTTTGAGTTAGTTAGCATTACCTGTTCCATCTTCT  
 CCTAGGAATGAGAGGAAGACTTAGCAGAACAAAGATACCATATGCTATAACATGCTAA  
 ACAGATGTGAGAAATACCATCTAACCTGGTGGTCCCAGCCGGCACTACAGGGAC  
 5658 TTAGCAGAACAGATATACCATATGCTATAACATGCTAAACAGATGTGAGAAATCACCA  
 TCTAACCTCCCTGGTTGGTCCAGGCCACTACAGGGACATTGGACTCTCTGGTGT  
 AAGTGAGATGGAGGAAGCCTGGTACAAGGCTGGTTAGCAATATTGACCAATTGGG  
 TTTCTTATTCTGAGTTCTACGTGCTGTATGACAATATTGACCAATTGGG  
 AAAAGCACCTTGAAAGCATAGATCATGGTTAGGTAGTGAGTGGTTTATTATTGTTGG  
 [A, T]  
 GAAGAGCCTTGGAGGTGCAGGGATCCATCCCCCTGGGGTCGGGAAGCATTCTGGGCC  
 TTTCTGGTTTCCATCGGTGTGGTCAAAACCTCTGATTTTGTGCTGGCTGGTGGGCC  
 CAGGTACCTTGCCTTACAGCATCTGGTGGTAACATCTGCTGAGCTGGGCCAG  
 AGTCGAAATTCCAGGTCTCAACAATGCCACCAATGAGAGCTATGTGGACACAGAGCCA  
 TGGAGGCTGAGAGGCTGACGTGTCAGCTAGCCTGCCTACTGCCATCATCCAGG  
 5945 ATTATTGTGTTGGAGAACAGCCTGGAGGTGCAGGGATCCATCCCCCTGGGGTCGGGAAG  
 CATTCCGGCCCTTCTGGTTCCATCGGTGTGGTCAAACCTCTGATTTTGTGCTGGC  
 TGGGTGGGGCACACAGGTACCTTGCCTTATCAGCATCTGGTGGTAACATCTGCT  
 GCAGCTGGCCCAGAGTCGAAATTCCAGGTCTCAACAATGCCACCAATGAGAGCTATG  
 GGACACAGCAGCCATGGAGGCTGAGAGGCTGACGTGTCAGCTAGCCTGCCTCAC  
 [T, C]  
 GCCATCATCCAGGTGAGGGGGCAGCCCCAACCCCTGCTAGAACGGCATCAGACCAACCTG  
 CCCCTCCCTCAAAGCCTAGCTTGATGCTAAATCTGATTAGGGGCTGGGTGGAGG  
 CTCTGCTGTAATCCAGCACTTGGGAGGTGAGGAGGGTGGATCACTTGAGGTGAGG  
 AGTTGAGACCACCTTGACCAACGTGATGAAACCCATCTCTACCAAAACACAAAAA  
 ATCCAGGCTTGGTAGTATGCCCTGAGTCCACCTACTCAGGAGGCTGAGGAGGAGA  
 6281 GCTAGAACGGGCATCAGACCAACCTGCCCTCCCAAAGCCTTAGCTTGATGCTAAATC  
 TGATTTAGGGGCTGGGTGTGGAGGCTCATGCCGTAAATCCAGCCTTGGGAGGCTGA  
 GGAGGGTGGATCACTTGAGGTGAGGAGTTGAGACCACCTTGACCAACGTGATGAAACCC  
 CATCTCTACCAAAACACAAAAATAATCCAGGCTTGGTAGTATGCCCTGAGTCCCACC  
 TACTCAGGAGGCTGAGGAGGAAATCACTGAAATCCGGGAGGAGAGGTTGAGTCAG  
 [C, T]  
 GAGATCGGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCGTCTAAAAAA  
 AAAAAAAAAAAAAAAAAAAACCAAGTTAGGGCTCACCTCTCCCTCCCCATCCAGG  
 GCTAAAGTGAACCTGAAAATTAAACAGTATCTCTCATCTGATGAGCAGGACCATACA  
 AAAAAACACAGCTGTAACCTGGTAAACTGTCCTGAGCTTAAACCTGAAAAGACTCAC  
 AGCCTCTCCATTATCCGTGGAGAACCCAACCTCTGCCAGCATAGCTTGAGACT  
 6452 ATGAAACCCATCTACCAAAATAACAAAATAATCCAGGCTTGGTAGTATGCCCTGT  
 AGTCCCACCTACTCAGGAGGCTGAGGAGGAAATCACTGAAATCCGGGAGGAGAGGTT  
 GCAGTGAGCTGAGATCGGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCGTCT  
 CAAAAAAAC  
 CCATCCCAGGGCTAAAGTGAACCTTGGAGAACCCAACCTCTGCCAGCATGAG  
 [G, C]  
 GACCATACAAAAACACAGCTGTAACCTGGTAAACTGTCCTGAGCTTAAACCTGAA  
 AAGACTCACAGCCTCTCCATTATCCGTGGAGAACCCAACCTCTGCCAGCATAGTC  
 TTGCAAGCTGTAATTCTCTAACATCCCTCACTCCGCTCCAGCCTCTGCTCCAAG  
 CCACAGCAGCAGTTGACAACATAAAATTGAGCTTCTGCAATGGTGCAGGAGGATTCTGC  
 TAGGTTTATGAAGGGAAAGCACAACATGACAGAACATGCAAGAGCAAAACACAGTCCAGAG  
 6610 GTGACAGAGCGAGACTCCGTCTAAAAAAAAAAAAAAAAACCAAGTTA

## FIGURE 3K

GGGCTCACCTCCTCCCTCCCCATCCCAGGGCTAAAGTGAACCTTAAAATTAAACAGT  
 ATCTCCTCATCTGCATGTAGCAGGACCATACAAAAAAACACAGCTGTACCTGGTTAAC  
 TGTCTGAGCTTAAACCTGTAAAAGACTCACAGCTCTCCATTATCCCAGGAGAAA  
 CCAACTCTCTGCCAGCATAGCTTGAGACTGCTAATTTCTAACATCCCTCACTCC  
 [T, G]  
 CTCAGCTCCTCTGCTCCAAGCCACAGCAGCAGTTGCACAACATAAATTGAGCTTCTGC  
 AAATGGTTGCAAAGGATTCTGCTAGGTTTATGAAGGGAGCACAACATGACAGAAATGCA  
 AGAGCAAACACAGTCCCAGAGAGCGCTTCTTCACTCATTCACTCATTGGTTTGCC  
 AAGAACTAGGCTAACCCCTGGGATACAAAGATAAGTAAGAAAGAGGTCATTACAAGT  
 TGCTCACAGCCCAGCAGAGGAAGGCCATGTCAACAGATAAATTGTATGCACTGAGAT

7247 GACACAGAGCAGAGTCACGGAGGACCTAAAGAGGAGGTGACACTCCACCTCTTTAAAG  
 GATGAGAACTTAACCAAGGAAACAAGGTATACAGAGGATGGTCAGGCAGAAGGGAAACAGTG  
 CCTAAAAACACTGAGGCTGAGAGAGTGTATCTGCGCAGGCAAAGTAAGGGGTTGGTG  
 TGGCTGGAGGGTAGAGGGCCAGAAGAGGATGGAAAAGTAGGCAGGAGCCAGACAATGAG  
 ATCTGGGTCTGTTCTGACAGCGACTTGGGCTGATTGGCAGTTATAAGGATCGTT  
 [T, C]  
 GGGCTACACAATGATGAGTGGGAGGTGGATTAGAATCAAGGCAGGGGACCTGTTGGGAGA  
 CTCTGCAGAGGCCAGGCAGGAATAATGCAAGCGAAGACAGGTAGAGAAAGAGATGGGG  
 CTGGACTTAAAAGAATGTTTACCAAGGAGCTGGTGTAGACTGGATGTGGGAGGTAAG  
 GGAGGATGACTCTCAAGTTTGGTGGCAACAGGTTAATGATGGTGTATTACTGA  
 GAGAGAAAACACTGGGGAGGACTAGACTTATTACAGATAAGCCAAAGCCAGAGGAGT

7360 AACAGTGCCTAAAACACTGAGGCTGAGAGAGTGTGATCTGCGCAGGCAAAGTAAGGGG  
 CTTGGTGTGGCTGGAGGGTAGAGGGCCAGAAGAGGATGGAAAAGTAGGCAGGAGCCAGA  
 CAATGAGATCTGGGCTCTGTTCTGACAGCGACTTGGGCTGATTGGCAGTTATAAG  
 GATCGTTGGGCTACACAATGATGAGTGGGAGGTGGATTAGAATCAAGGCAGGGACCTG  
 TTGGGAGACTCTGCAGAGGCCAGGCAGGAATAATGCAAGCGAAGACAGGTAGAGAAAAG  
 [A, G]  
 GATGGGGCTGACTTAAAAGAATGTTTACCAAGGAGCTGGTGTAGACTGGATGTGGG  
 AGGTAAAGGGAGGATGACTCTCAAGTTTGGTGGCAACAGGTTAATGATGGTGTCT  
 TTACTGAGAGAAAACACTGGGGAGGACTAGACTTATTACAGATAAGCCAAAGCCA  
 GAGAGGATGATGTGACAGAAAGGCCATGCTCAAAGGAGCTGAAGGCTGATGGCAGCCA  
 TGTAGAGCACAGTGAAGGCAGGTGAAGGTACAGATGGCCAATTCCCTCAAGCTACTG

7644 GACCAGGTAGAGAAAGAGATGGGCTGGACTTGAAAGAATGTTTACCAAGGAGCTGGT  
 GATGACTGGATGTGGAGGTAAAGGGAGGATGACTCTCAAGTTTGGGGCAACCAG  
 GTTAATGATGGTGTCTTACTGAGAGAGAAAACACTGGGGGAGGACTAGACTTATT  
 CAGATAAGCCAAAGCCAGAGAGGATGATGTGACAGAAAGGCCATGCTCAAAGGAGCTGA  
 AGGTCTGATGGCAGCATGTAGACAGCACAGTGAAGGGCAGGTGAAGGTACAGATGGTCCA  
 [A, T]  
 TTCCCTCAAGCTACTGCTACGCTAGGACTGCACGGAGCTCCAGACCTGCGTGTGTGGG  
 GCAGGGCTGTTGGAACTGCTGAACCACATTGGTCTTCCGCCACCAACCACCCCTTCTCC  
 TCTCAGATGGTCTGGCTTCTGAGCTTGGCTTGTGCCATCTACCTCTCCGAGTCC  
 TTCTCCGGGCTTCTGACGGCCGCCGCCCTGAGCTCTGATTTGGTGTCAAGTAC  
 ATCTTCGGACTGACCATCCCCCTTACAGGCCAGGGTCCATGCTTTGTGAGTCTG

8127 CATCCGGGGCTTCTGACGGCCGCCGGCTGCAGATCCTGATTCGGTGTCAAGTACAT  
 CTTCGGACTGACCATCCCCCTTACACAGGCCAGGGTCCATCGTCTTGTGAGTCTGGG  
 GATGCACCCCTGCCATTGGAGCAAGGCTCCAGCAGACATGAGGAGGATGACTGTTT  
 AAGATGTCGTGAGCTCTCATTGCAAGGGCTGGTTAGCTGTTGAGAGGAGATTCTG  
 AGGGGGTTCTGTCTGGGAGGGTCAAAGTCATGACTCACAGAGGTTCTGGTAGTTAAT  
 [A, C]  
 CCTGCAGAAAAGAGCTGTACATTCTGCCAGTTCCCCATTCTAGTGCCTCAACCCCTCC  
 CTGCCTGGAAAAGTCTGCCATTGCTAACTCCATCCCCCTCCCTCAGCCCAAACCTCT  
 TCTAAAGAAAAGAAAGCATTCTTCTAGACAAGTCCCCATGTCCTTGGAAA  
 GGGCGGTGGCGACGGGACAGGGTCTGATCAGGGTTTAAATTCTGTCTGGTGTGCT  
 CCATTAGCTTGTGATGGCATCCCTCCCTGGGTCAAGACACCCAAAGTGGGTATTATGGG

8317 GAGCTCCTCATGCAAGGGCTGGTTAGCTGTTGTCAGAGAGGATTCTGAGGGGTTTC  
 TGCTTGGGAGGGTCAAAGTCATGACTCACAGAGGTTCTGGTAGTTAATACCTGAGAA  
 AAGAGCTGTACATTCTGCCAGTTCCCCATTCTAGTGCCTCAACCCCTCCCTGCC  
 AAGTCCTGCCATTGCTAACTCCATCCCCCTCCTCAGCCCAAACCTCTAAAGAA  
 AAAGAAAGCATTCTTCTAGCACAAGTCCCCATGTCCTTGGGAAAGGGCGGTGG  
 [G, A]  
 CGACGGGACAGGGTCTGATCAGGGTTTAATTCTGTCTGGTGTGCTCCATTAGCTT  
 TGATGGCATCCCTCCCTGGGTCAAGACACCCAAAGTGGGTATTATGGGAAAGAGGGT  
 GGGAGCCTGTGAGCATGCTTTCCCCAGACCTTCATTGACATTGCAAAACCTC

## FIGURE 3L

|       |   |
|-------|---|
|       | CCCCACACCAACATGCCCTCGCTCATCTCGCTCTCATCAGCGGTGCCTCCCTGGTGTG<br>GTGAAGGAGCTCAATGCTCGTACATGCACAAGATTGCTTCCCATCCCTACAGAGATG   |
| 9079  | TTCTACTGCTCTAAATAATTCCCCCTAAGGAGGCAGGGGAGTGGGATTCAAGGTCCCCAGA<br>GAAAAGGGAGACTTGAGAGAGACGCCCTGCCCTGGCCCCACCTTAGGGCAATCCCCATT<br>TCCACTCTGGGTTTGAGGTGGTGGCAACAGTATCTCCGGGGCTGTAAGATGCC<br>CAAAAAGTATCACATGCAGATCGTGGGAGAAATCCAACGCCGTGAGTCCAGGTGGCCA<br>GAAGCCTGCCACCCGACCTCATGCCCACTAAGGCCTGAGCTGGAGAGGGAGACAA<br>[G, A]<br>ATGAACTCTATGAAAGTGCAGTCAGAACTGTATGACACTGACCATGTATGAATTATTACT<br>ATTACCGTTCTGAGAAGGGCCACAACCAAGCCAATGTAGGCTATTTATGAGAAATG<br>AGTCTTAACGCTGCCACACTCCCCCTATAAACTCATTCACACTGATGCTTAAACAAAGCC<br>TCTCTGAACAGCCGCTGGCTCTTGCTTAATGCAATTGGTTCTTGCCAT<br>GTAGAAAGGAACTATTAGGTTAACCGAGATTCAAGGCATCCACTCTGTGCCAGGCAC |
| 9537  | AACTGATGCTGTTAAACAAAGCCTCTCTGAACAGCCGCTGGCTCTTGCTTC<br>TAATGCATTGGTTCTTGTCCATGTAGAAAGGAACTATTAGGTTAACCGAGATTCA<br>AGCATCCACTCTGTCAGGCACCATGCTGGCCCTGGGAGGAGAGGGGTGACGCTTGTC<br>CTGCAAGGGTTGAAACAGGCAAGGGAGGGAGACACATAGCACCAAAAGGTCTAGGGTCT<br>GTGGACTCGTGAGCATAAGGGTTCAGAACTCTGGAGTTAACAAACGAGGCCCTACCACA<br>[G, T]<br>ACTGGCCCCGGGACCTGGGCAAGTTAGGTTCTCTCAGCCTCAGTTCCCTTTGTAAA<br>ACAGGAGTGTGGTCCCTACCCATGGGCTGAGGATTCAAGACTGGATGGATAA<br>CTTAGGCAAGATCCGGCACACCATGGGGCTGGCTGGCTGGCTGGTGAAGG<br>ACTTGGCTGCCCTCCCACTCACACCCCTGGGTTCTGCCCTGGCTCTCGCAG<br>GTTCCCCACCCGGTGTGCCCTGGTCACTGAGGACATGAGGACAGCCTT                              |
| 12302 | AGCCCCACATAACCTATGGGAGAGGATTACTAAACTTTCTAACGGTGTGAAACCAA<br>GGCTCAGAATGGTAAGTAAATTGTCACAGGGCACAGAGGTAGGGTAGGCTCTGG<br>ATTAAAACTCCAAGTCCTGGACTCCAGACCTCTAGGCTGACTGTCTCATAGGGAGGA<br>GTCTCACCCACCTAGGGCAGAGAAAATCTTAAAGCCAGAGAAGTGAAGTGGCTCATC<br>TGTGGTACCCAGAGAGACAGTGTGAGGACAGGGAGAAAATTATAACCTCAGTTCCAG<br>[C, G]<br>CCAAGGACTGCTTTGACCATAACCCAAACAAGCCCCGCTATGGTGTATTCTTAGGT<br>TCATATGGCGGCTTTGTTCCATTGATCTTCACAGCAATTCTCATACAGGAATCTGGC<br>AGATTTATTCTCTTAGAGGAATTCCAGGTCTAAATCTATAGGGGCAACTATCAA<br>ACTTCACCCAAATGTTGGCCCTACCCACACACAAAAGGGCCCCAGCCGATCAGAAAG<br>CACTGCTGAGCTCTGTCAGGGCCCACGCACTGCTGTGAGACAGAGAGGGAACTCA        |
| 12354 | GAAACCAAGGCTCAGAATGGTAAGTAAATTGTCACAGGCCACAGAGGTAGGGAGGTG<br>GAGTCTGGATTAAAACCTCAAGTCTGGACTCCAGACCTCTAGGCTGACTGTCTCATAG<br>GGAAGGCACTCTCACCCACCTAGGGCAGAGAAAATCTTAAAGCCAGAGAAGTGA<br>GGCTCATCTGTCAGGACAGTGTGAGGACAGGGAGAAAATTATAACCTCA<br>GTTCCCAGCCAAGGATCTGTTGACCTAACCAACAAAGCCCCGCTATGGTGTATT<br>[C, T]<br>CCTTAGGTTCATATGGGGCTTTGTTCCATTGATCTTCACAGCAATTCTACAGGA<br>ATCTGGCAGATTATTCTCTTAGAGGAATTCCAGGTCTAAATCTATAGGGGCAA<br>CTATCAAACCTCACCAATGTTGCCCTACCCACACACAAAAGGGCCCCAGCCG<br>TCAGAAAGCACTGCTGAGCTCTGTCAGGGCCCACGCACTGCTGTGAGACAGAGAG<br>GGAACCTACATTATTGATCACCTACTGAGCATCATCAGAGCTAGGCTAGGACCGTCACATT                     |
| 12487 | ACCCACCTAGGGCAGAGAAGAAAATCTTAAAGCCAGAGAAGTGAAGTGGCTCATCTGTGG<br>TCACCCAGAGAGACAGTGTGAGGACAGGGAGAAAATTATAACCTCAGTTCCAGCCAA<br>GGATCTGCTTGACCTAACCAACAGCCCCGCTATGGTGTATTCTCTAGGTTCAT<br>ATGGCGGCTTTGTTCCATTGATCTTCACAGCAATTCTCATACAGGAATCTGGCAGAT<br>TTATTCTCTTAGAGGAATTCCAGGTCTAAATCTATAGGGGCAACTATCAAACCT<br>[C, T]<br>ACCCAATGTTGCCCTACCCACACACAAAACAGGGCCCCAGCCGATCAGAAAGCA<br>CTGAGCTCTGTCAGGGCCCACGCACTGCTGTGAGACAGAGAGAGGGAACTCACATT<br>ATTGATCACCTACTGAGCATCATCAGAGCTAGGACCGTCACATTCTTAACTTTGA<br>ATCCCTTCATGAGGTAGGCATTATTCTCTTTGTTACATAGCCATTAAAGAAC<br>AAATTGGGGCTGGGTGACTCACACCTGTGATCTAGCACTTAAAGGGCTGAGGC                 |
| 13198 | CTAACTATTCAAGGAAGGTTAGGCCGGAGCACAACATTGGGTTCCAGGGTTGAGGCTCCAG<br>TGAGCTGATCTGCCACTGCACTACAGCCTGAGCAACAGAGCAAGACCCCTGTGACTCCAA<br>AAACAAACAAACACACATTGAAACCAACAGATCTGACCCAAAGATGCACTGCTTAA<br>TAGATGCCACCTCCCTGTGCTGGGCTTCACTAAAACACAGACAGATCAGGCAAC<br>CAAGTCATCTAAGGAAAGAGGAAAGTGTAAACCAAAAGCACAAATACATAAAATTG   |

## FIGURE 3M

|       |  |   |
|-------|--|---|
|       |  | [ - , A ]   |
|       |  | AAAATGCTATTTAAAGAAAAAAAAGAGAAGAGAGGGCTTGAGGTTGACTAACAGAGAAAT<br>GGCTTGGCTAATCCAGGAAGACTCCTGAAAGAGGTTTTTTCCCCAGGTCTGCTT<br>TGACATCTCTTTCACAGTGATCTGGTAGTGAGCTCCTCTCCCTCTCC<br>GCCTGCCCTATGGTGGCAGTGGGTGCGCTCTCCGTCTGGTGTGGCTTCCAGA<br>CTCAGTTGAAGTGATAGCTCCGCCCTCTAGGCCACAGTCGGTCCCTGGGCCAGCC  |
| 13257 |  | GTGAGCTGATCTGCCACTGCACTACAGCCTGAGCAACAGAGCAAGACCCCTGTGACTCCA<br>AAAACAAAACAAAACACATTGAAACCCAAACAGATCTGACCAAGATGATGATGCTCTT<br>ATAGATGCCACCTCCCTGTGCTGGGCTTACTAAAAACACAGACAAGATCAGGAA<br>CCACAGTCATCTAAGGGAAAGAGGAAAGTGTAACCAAGCACAAATACATAAAATATTG<br>CAAAAATGCTATTTAAAGAAAAAAAGAGAAGAGAGGGCTTGAGGTTGACTAACAGAGA<br>[ A , G ]<br>TGGCTTGGCTAATCCAGGAAGACTCCTGAAAGAGGTTTTTTCCCCAGGTCTGCTT<br>TTGACATCTCTCTTTCACAGTGATCTGGTAGTGAGCTCCTCTCTCC<br>AGCCTGCCCTATGGTGGCAGTGGGTGCGCTCTCCGTCTGGTGTGGCTTCCAG<br>ACTCAGTTGAAGTGATAGCTCCGCCCTCTAGGCCACAGTCGGTCCCTGGGCCAGC<br>CGCAAAGGGCTTCATGCCACGGCTGGCTAGTCCACTGTACCTCCACCTCTGGGCC                                |
| 14541 |  | TCATGGACACTGACATTATGTGAATCCAAGACCTATAATAGGGTAGGTAAATTCAAGCT<br>TATGACCTCTTCTTTGCTCTGACCCACCCAAAGAAGAGGTTGCTTTTTAAAGCCAATA<br>AAGACATTCTGCAACTTGAGCTCAGTCCCTGTACAGGCCAGGATATCCAGGGAT<br>TAAAATCATCACGTACTGCTCCCTCTACTTTGCAACTCAGAGATCTTCAGGCAAAA<br>GGTCAATGCCAAGGTAAAGGTCAGTCCCTGGCGACCAGAGGCTCTGGACAGAGTGGCC<br>[ G , A ]<br>GAAAATGGAAGCAGAAGGGCGGTGGGAGCTGAGAATAGGCCACTCCATAGAGGGTGGAG<br>GTCAAGATTGCTTGGCTCTCCCTGCAAGACAGGCATGGACCCCAGAAAGTATTACT<br>AGCCAAGAAAAATACCTCAAGAACAGGAGAAGCGGAGAATGAGGCCACACAACAGAG<br>GAGGTCTTATTCTGAAACCAAGGTGAATGAAGGCCAGAAGCAGCCCCGTGCCCTGCT<br>CTCCGCCATTCTGATACTGCCCCCTGTTACTCATGGTACCCCTGGGGGCCCCGCTTCCC<br>14545 |
|       |  | GGACACTGACATTATGTGAATCCAAGACCTATAATAGGGTAGGTAAATTCAAGCTTATG<br>ACCTCCTTCTTTGCTCTGACCCACCCAAAGAAGAGGTTGCTTTTTAAAGCCAATAAAGA<br>CATTCTGCAACTTGAGCTCAGTCCCTGTACAGGCCAGGATATCCAGGGGTTAAA<br>ATCATCACGTACTGCTCCCTCTACTTTGCAACTCAGAGATCTTCAGGCAAAGGTC<br>ATGCCAAGGTAAAGGTCAGTCCCTGGCGACCAGAGGCTCTGGACAGAGTGGCCGAA<br>[ A , G ]<br>ATGGAAGCAGAAGGGCGGTGGGAGCTGAGAATAGGCCACTCCATAGAGGGTGGAGGTCA<br>AGATTGCTTGGCTCTCCCTGCAAGACAGGCATGGACCCCAGAAAGTATTACTAGCC<br>AAGCAAAATACCTCAAGAACAGGAGAAGCGGAGAATGAGGCCACACAACAGAGGAGG<br>TCTCTATTCTGAAACCAAGGTGAATGAAGGCCAGAAGCAGCCCCGTGCCCTGCTCTCC<br>TGCCCATCTGATACTGCCCCCTGTTACTCATGGTACCCCTGGGGGCCCCGCTTCCCACCC<br>15041 |
|       |  | ACCAAGGTGAATGAAGGCCAGAAGCAGCCCCGTGCCCTGCTCTCTGCCATTCTGATAC<br>TGCCCCCTGTTACTCATGGTACCCCTGGGGCCCCGTTCCACCTGACAGGCAAAGACA<br>GAAAGTCTGGAAACACTGCCGGTGGCCGCTGGCATTTTCTTTCTTTCTTTCTTT<br>TCTTTAGAGATGGAATTTGCTTGTCACTGGCTTGTGAGTGCAATGGCGTTATCTT<br>GGCTCACTGCAACCTCCACCTCTGGGTTCAAGCGATTCTCTGCCTAGCCTCCAAAGT<br>[ C , A ]<br>GCTGAGATTACAGGTGCCACCAACCCCAGCTAATTTTGTTAGTAGATATTGGGT<br>TTCACTGGTGGCCAGGCTGGTGTAAACTCCTGACCTCAGGTGATCCACCTACCTTAG<br>CCTTCAAAGTGTGGGATTACAAGCCTGAGCCACTGCCAGCCTGGCATTTTCTT<br>CTTGGATGAGGTGTACCATCTCCAGGAAAGCCACTGAACCCCCAAGGCCCTTCCAT<br>TTCTGGCTAAGATAGACATGGACTTTGAACAACCCAGAGGGGAAACAGCA<br>15053                       |
|       |  | GAAGGCCAGAAGCAGCCCCGTGCCCTGCTCTCTGCCATTCTGATACTGCCCTGTTA<br>CTGATGGTACCCCTGGGGCCCCGTTCCACCTGACAGGCAAAGACAGAAAGTCTGG<br>GAACACTGCCGGTGGCCGCTGGCATTTTCTTTCTTTCTTTCTTTCTTTAGAGA<br>TGGAAATTGCTTGTCACTGGCTTGTGAGTGCAATGGGTTATCTGGCTACTGCAA<br>CCTCCACCTCTGGGTTCAAGCGATTCTCTGCCTAGCCTCCAAAGTCGCTGAGATTAC<br>[ A , C ]<br>GGTGCCACACACCCAGCTAATTTTGTTAGTAGATATTGGGTTTACCATGTTG<br>GCCAGGCTGGTGTCAAACCTCTGACCTCAGGTGATCCACCTACCTAGCCTCCAAAGTG<br>CTGGGATTACAAGCCTGAGCCACTGCCAGCCTGGCATTCTCTGGATGAGGT<br>GCTACCATCTCCAGGAAAGCCACTGAACCCCCAAGGCCCTTCCATTTCTGGCTAAG<br>ATAGGACATGGCCATGGACTTTGAACAACCCAGAGGGGAAACAGCAGTGAAATTCTG<br>15065                         |
|       |  | CAGCCCCGTGCCCTGCTCTGCCATTCTGATACTGCCCTGTTACTCATGGTACCC  |

## FIGURE 3N

TGGGGGCCCGCTTCCACCCGACAGGCAAAGACAGAAAGTCTCTGGGAACACTGCCTG  
 GTGGCCGCTGGCATTTCCTTTCTTTCTTTCTTTAGAGATGGAATTTGCT  
 CTTGTCACCCAGGCTTAGTGAATGGCCTTATCTGGCTCACTGCAACCTCCACCTCTG  
 GGTTCAAGCGATTCTCTGCCTTAGCCTCCAAAGTCGCTGAGATTACAGGTGCCACAC  
 [A, G]  
 CCCAGCTAATTTCGTATTTAGTAGATATTGGTTTACCATGTTGGCCAGGCTGGTG  
 TCAAACCTCTGACCTCAGGTGATCCACCTACCTAGCCTCCAAAGTGTGGGATTACAA  
 GCCTGAGCCACTGCGCCAGCCTGGGCTTTCTCCATTTCCTGGATGAGGTGCTACCATCTCC  
 CAGGGAAAGCCACTGAACCCCCAAGGCCCTCTCCATTTCCTGGTAAGATAGGACATGGC  
 CCATGGACTTTGAACAACCCAGAGGGGAAACAGCAGTGAATTCTGGGAACCCAGGC

15108 TGTTACTCATGGTACCTGGGGGCCGCTTCCCACCCGACAGGCAAAGACAGAAAGTC  
 TCTGGGAACACTGCCTGGTGGCCGCTGGCATTTCCTTTCTTTCTTTCTTTCTTT  
 AGAGATGGAATTTGCTCTTGTACCCAGGCTTAGTGAATGGCCTTATCTGGCTCAC  
 TGCAACCTCCACCTCTGGGTTCAAGCGATTCTCTGCCTAGCCTCCAAAGTCGCTGAG  
 ATTACAGGTGCCACACACCCAGCTAATTTCGTATTTAGTAGATATTGGTTTAC  
 [A, C]  
 TGTTGGCCAGGCTGGTGTCAAACCTCTGACCTCAGGTGATCCACCTACCTTAGCCTCCA  
 AAGTGTGGATTACAAGCTGAGCCACTGCGCCAGCCTGGCATTTCCTGGAT  
 GAGGTGCTACCATCTCCAGGGAAAGCCACTGAACCCCCAAGGCCCTCTCCATTTCCTGG  
 CTAAGATAGGACATGGCCATGGACTTTGAACAACCCAGAGGGGAAACAGCAGTGAATT  
 TCCTGGGAACCCAGGCCAGGGCTAGCAAGGCTGGGTGGCATGGCAGTAATCCT

16274 CTTCCAGACTGTCTCCCTGCAGGAGCTGCAGCAGGACTTGTAGAATGCGCCCCCACC  
 CCCCCAACACAACACCAGACCCCGCTAACGGCACCGCGTGTCTTATACACCTTCAGCCC  
 TGACAGCTCTCACCTGCCAGAGTGAGCCACCAGCCTCGTGGGCCCCGGCGAGCC  
 CAGTGACATGCTGGCCAGCGTCCCACCCCTCGTACCTTCCACACCCATCTGGACAT  
 GAGTGGAGTCAGCTCGTGGACTTGTGGCATCAAGGCCCTGGCCAAGGTGAGGCCCTC  
 [-, G]  
 GGGACAGCAAGCACCACCCACTCCACCCCTCCGCTCTCCACATTCCCTTCTG  
 GGAGCCCTCATTCAGGAAGCTGAGGGAGGAAGCTACTGGGAGACTAACAGCTCTAG  
 GAATCCCTCTTCCCCAGGCCACAGGTTGAGACATTCTCCACAGAGCAGGGCCAGA  
 CGGCCCATGACAATGAGTGGCGGACAAGTCTACCAGAGTTCAAGGCCCTGTGCTCCCA  
 ACACCCCCAGCAGTGGCATCCAAAGTCCCTCAGCCATCAGGAACCCACCCAGGTTCT

17424 AACATGGTAAACCCGCCTCTACTAAAAATACAAAATTAGCCAGGTGTGGTACGGGC  
 CCCTGTAGTCCCAGCTACTCGTAGGCTGAGGCAAGAATTGCTTGAACCCAGGAGGCGG  
 AGGTTGAGGCAAGATCGGCCACTGCACTCCAGCCTGGCAACAGAGTGAACACT  
 CATCTAAAGAAAAAAAGAAAATATCTAGCCCCAACAGAAGGGCCATGGTACTTT  
 AAGTGGCCGCACGTTGCAAAGTCCATTCCGCTCCACTTCCAGAGAAACCGTCAGC  
 [C, T]  
 AACACTCCAGGGAGAAGTGGTGTCTTGTCTATTTCGTCTGGCTGCTGGCTCT  
 CAGGGTTGCTTATTTGTTGGCTTCCCTCTGAAGTACGTTGTGAATCACTTTGAGA  
 CCCACTGACAACATTCTTCTTGTGCTCCCTACCCAACACACTCTAGCTGAGCT  
 CCACCTATGGGAAGATGGCGTGAAGGTCTTGTGAACATCCATGGTAAGAGAAAGA  
 GGACATTAGGGACTGAAAGACTGGCAAGGAGTGTGGGTAGGAACAGGTTGGGGTC

17627 AATATCTAGCCCCACAAGAAGGGCCATGGTACTTAAAGTGCCGCCACGTTGGCAAA  
 GTCCATTCGCTCCACTTCCAGAGAAACCGTCAGCCACACTCCAGGGAGAAGTGGT  
 TGCTTGTGCTGATTTGCTTGGCTGCTGGCTCTCAGGGTTGCTTATTGTTGGC  
 TTCCCTCTGAAGTACGTTGTGAATCACTTGTAGGCTGAGCTCCACCTATGGGAAGATCGCGT  
 [G, A]  
 AAGGTCTTCTGGTGAACATCCATGGTAAGAGAAGAGGACATTAGGGACTGAAAGACT  
 GGCAAGGGAGTGTGGGTAGGAACAGGTTGGTGGGCTGTAATAGTGGAGGAGGTTGGAAAC  
 GAGAGCACCAGCTATCCCCACAAGCTGTCGCTGCTCATAAAGCTTCAGGTACAAGT  
 CCAAAGAGACTGGTCAGATTGATAAACATCTAGGGCCTTAGTGAAGAGTGGGGTG  
 AGGAGGTATGGAGTTACAGAAGGACAGCTAGGATTCTAATCTACCCATAACTAATTG

18427 GGGTGCATATACACAGCCTCAAGGACGTGGCACAGGGCAGCAGACATTACATGACTG  
 CATGTACGAAAGTGCAGAGATGTGGAGCAAGTGCACACAGACACAGGAGAATGTGA  
 AGGGGCACATACACACACACACCCAGCTCCCTGCACTGGCTGAGACCCCTCCAGCAGGGCT  
 GCAGTTCCAAGCTCCGATGGCACGTTGGGAGAGAATCTGCACTGGCAATGACCTG  
 CTATGATATGTTCTGGAGTTAGAAGCAGTGGATTCTCCAAACCTCACTGGACACCCCT  
 [T, C]  
 AGGAAACCATCTCTAGGATTAAGAGTAATCCACACAAACTTCAATGCCACACATTGGAA  
 GTTGCTGAAAGTCTGGAAAACAAGAGGAAGGATGGGCTTGGGGATAGAAGTGGC  
 AGCGGCCTTCAAGGATGGCTTAGGCTTCCACTCGAATCACCACAAAGTACTGACTC

## FIGURE 30

CCTAAATCAAACGTCTCCTCTGCTCTGGGTGAAACTTCAGCATCCTCAAGTCATGT  
 TGCCCTGCGTCCAGAACTGATATTGCACTGCCATGCCATGGCCTCAGATACAGCA

18813 AGAGGAAGGATGGGTCTTGGGGATAGAACTGGCAGGCCCTTCAGGATGGCTAG  
 GCTTTTCACTCGAACATCACCACAAAGTACTGACTCCCTAACTCAAACGTCTCCTGC  
 TCTGGGTGAAACTTCAGCATCCTCAAGTCATGTTGCCCTGCCGTCCAGAACTGATA  
 TTGCACTGCCATGCCATGCCCTCAGATACAGCAAGAGCTGGACCTCAGGCCCTCCC  
 ATCCCTGCTCTGGTCTACTATCTCCCCACCCCCAGCTCAATCCACAATGGCTTAT  
 [C, G]  
 TTTCTGAAGGTGATCTTTCTCCTCTAGCCAGGTGACAATGACATTAGCCATGGAGG  
 CGTCTTGAGGATGGGAGTCTAGAAATGCAAGCACGTCTTCCCAGCATACATGACCGAGT  
 CCTCTTGCCAGGCAATGCTAGAGACGTGACCCAGGACACAACCTCCAAGGGTAAG  
 GTTCTTGACACCTGGGAATCCTAGGCTCAAGGCACTGAAATAGCAGGACCAAGGGCAT  
 TATTAGAAGAACACAGGAGAAGGTTAAGTCCAATCAAGTCTGCCATTCAGTTT

19035 GGACCTCAGGCCCTCCCATCCCTGCTCTGGTCACTATCTTCCCCACCCCCAGCTCCA  
 ATCCACAATGGCTTATCTTCAGGATCTGAAAGCTGATCTTCTCTAGCCAGGTGTACA  
 ATGACATTAGCCATGGAGGCCTTGGAGGATGGGAGTCTAGAAATGCAAGCACGTCTTC  
 CCAGCATACATGACGCAGTCCTTCCCAGGCAATGCTAGAGACGTGACCCAGGAC  
 ACAACTTCCAAGGGTAAGGTTCTGCACCTGGGAATCTAGGCTCAAGGCACTGAAA  
 [T, C]  
 AGCAGGACCAAGAGGCATTAGAAAGAACACAGGAGAAGGTTAAGTCCAATATCAA  
 GTCTGCCATTTCAGTTCTGAATCTGTTCTTATCTATAGAATGAGCACCCTCACTA  
 ACATTACCTACCTCTGCACTTTCTTTTGTATTGTTAGGGTTAAATGATAATTACA  
 TCTTTGTGCACTTGAAGACCTTGTGATTGTTAAATCTTATCAATATAAGTT  
 TCTGGTTGCACAAACACCCAAAGCATAGTAGAGCAGGCCACTCTGCTGGCATGTTCC

19182 AGGATGGGAGTCTAGAAATGCAAGCACGTCTTCCCAGCATACATGACCGAGTCCTCTTG  
 CCCAGGCAATGCTAGAGACGTGACCCAGGACACAACCTCCAAGGGTAAGGTTCTTG  
 ACCTGGGAATCCTAGGCTCCAAGGCACTGAAATAGCAGGACCAAGAGGCATTATTAGAA  
 AGAACACAGGAGAAGGTTAAGTCCAATATCAAGTCTGCCATTTCAGTTCTGAATCT  
 GTTCTTATCTATAGAATGAGCACCCTCAACTAACATTACCTCTGCACTTTTC  
 [T, C]  
 TTTATTTGTAGGGTTAAATGATAATTACATCTTGTGCACTGAAAGCACTTTG  
 TGTTAGGTTAAATCTTATCAATATAAGTTCTGTTGACAAACACCCAAAGCATA  
 GTAGAGCAGGCCACTCTGCTGGCATGTTCCCTGCCTCCTCATCTCTTCTAAAGG  
 GGGCTTCGGGAAGGGAGGGAGGAGTAAGCCTACCCATTAACTTACCGGAGCTTAA  
 GAGATTCAGGCTGGTGGGATAAAGAGATTGGGCTGAGTTGTCTAGCTTGTGTTGA  
 19508 TAATTACATCTTGTGCACTGAAAGCACTTGTGATTGTTAAATCTTATCAAT  
 ATAAGTTTCTGGTTGACAAACACCCAAAGCATAGTAGAGCAGGCCACTCTGCTGGCA  
 TCGTTCCCTGCCCTCTCATCTCTTCTAAAGGGGGCTTCGGGAAGGGAGGGAGGG  
 GAGTAAGCCTACCCATTAACTTACCGGAGCTAGAGATTCTAGGCTGGTGGGATAAA  
 AGAGATTGGGCTGAGTTGTCTCAGTTTGTGACATTAACTTACTAGCTCAGTAAGT  
 [-, G, C]  
 ATACAAATGGGATACAAATAACACCATCTAAACACTCCAGAAGACTGGGAGTCAGAAAAAA  
 TCCTACCTCTGGGGCTGCCAGATCCCCAGTCATCTAGCCCTCAGGGTCCCT  
 CCCAGCTCAGCTCTGCCCTGGGCTCCAAAGACTCTTGTGCCCCAGCCCTGGTAA  
 AACACCCCTGCCCTGTGGGCTAAGAAAGGTTTCTGGCCCTAGAGCAATGATT  
 TGCTCTTGCCTTAAGAGACTGATGAAGGTGAAACCATCTGTTCAAGTGTGAAAGACT

19571 AGTTTCTGGTGACAAACACCCAAAGCATAGTAGAGCAGGCCACTCTGCTGGCATCG  
 TCCCCTGCCCTCTCATCTCTTCTAAAGGGGGCTTCGGGAAGGGAGGGAGGGAG  
 TAAGCCTACCCATTAACTTACCGGAGCTAGAGATTCTAGGCTGGTGGGATAAAAGA  
 GATTGGGCTGAGTTGTCTCAGCTTTGTGACATTAACTTACTAGCTCAGTAAGTCAT  
 ACAAAATGGGATACAAATAACACCATCTAAACACTCCAGAAGACTGGGAGTCAGAAAAATC  
 [T, G, C]  
 TACCTCCTGGGGCTGCCAGATCCCCAGTCATCTAGCCCTCAGGGTCCCTCCC  
 AGCTCAGCTCTGCCCTGGGCTCCAAAGACTCTTGTGCCCCAGCCCTGGTAAAAAA  
 CCTCCCCCTGCCCTGTGGGCTAAGAAAGGTTTCTGGCCCTAGAGCAATGATT  
 TCTTGTGCTTAAGAGACTGATGAAGGTGAAACCATCTGTTCTAAGTGTGAAAGACTGCC  
 CAGGAACACACAGGGCGCTGGCTCTGCCCTCATGCCCTAGAGGGAAACCTGGGAAAC

20147 GCCTAGAGGGAAACCTGGGAAACAACGGGCTTCTGCTCGTGAATTTGTCCGCAG  
 AGCAAAGAGGGAGATTCTGGAGGAAGCTGCATTAGTTGTAGTGCCTAAATCATGTTAG  
 CTACTCTAGTTGGTATGATACTTGTAGTAGTCATAGCACTTATAATAATTATTTA  
 TATAATATATACTTACATTAGACCACTCACAGATACAAATCACACACATAAACACA  
 CACCTTCAACAGCATTGTGAGGGACAAAGCAGGCAAAGTGGCTGGTATCAGACTT

|       |  |  |
|-------|--|--|
|       |  | [T, G]<br>AACAGATTAGAAAATATTCAGGAGGACAGGAATTCCCAAGGTCAAGGCAGCTAGCC<br>AATAGTTTCTAAGCTGAGTAAACCTTCCCTGCCTAACGGCCACAAAGGAGGGAA<br>GACCGCGATACACACCTGTCTGGTATAAGGGGAAGACACAGCCGTGCTGTTTGTGA<br>GGCAGGTAAAGGAAGGGCAAGAGGATAAGTCATGTGTCAGGAAGCAGCGTCAACCAGA<br>GCCGGCCACCTGTCCCTTCTGCCACCATGCACCAACTTGCTGTTAGTCAGTCACTGAAG  |
| 20180 |  | TTCTGCTTCGTGAAATTGTCCAGAGCAAAAGAGGGAGATTCTGGAGGAAGCTGCATT<br>AGTTGTTAGTGCCTAATCATGTCAGTCACTCTAGTTGTATGTACTTGATTAGTC<br>TAGCACTTATAAAATAATTATATTATATAATATACATACATATTAGACCAATTCA<br>CAGATACAAATCACACACATAAACACACACCTTCAACAGCATTGAGGGACAAAGCA<br>GGCAAAAGTGGCTGGTTACAGACTTAAACAGATTAGAAAATATTCAGGAGGACA<br>[G, A]<br>GAATTCCCCAAGGTAGGCAGCTAGCCAATAGTTTCTAAGCTGAGTAAAACCTTCCCT<br>GCCTCTAACGGCCACAAAGGAGGGAAAGACCCGCGATACACACCTGTCTGGTATAAGGGG<br>AAAGACACAGCGTGTGTTTGTAGGGCAGGTAAAGGAAGGGCAAGAGGATAAGTC<br>TGTGTCAGGAAGCAGCGTCAACCCAGAGCCGCCACCTGTCCCTTCTGCCACCATGC<br>ACCAACTTTGCTGTTAGTCAGTCACTGAAGCTTCTGCACTGGCTTCCCTCCAGGCT  |
| 20584 |  | TGTCTGGTATAAGGGGAAGACACAGCCGTGCTGTTTGTGAGGCAGGTAAAGGAAGG<br>GGCAAGAGGATAAGTCATGTCAGGAAGCAGCGTCAACCAGAGCCGCCACCTGTCCC<br>TTTCTGCCACCATGCACCAACTTGTGTTAGTCAGTCAAGCTCATTCTGCACTGGC<br>TTCTCCCTCCAGGCTCCAGGGGATGCTGAGCTCTCTGTACGACTCAGAGGAGGACA<br>TTCGAGCTACTGGGACTTAGAGCAGGTGAGCTGAGGGAGGGCTGTGAGGGTGGAGC<br>[A, T]<br>GGGCGAAGAGGGGAAGGGATGGGTCGCTGTCAAATACAAGGCCTTCACTCAGCTGTC<br>CCTCCAGCCCAGAGCAGTCACATTCAAGGCCACAAAGATTGTGGTCACTTTGTTT<br>TTCTTTCTTTCTTTTTTTTTAATTGAGACAAAGTCTCACTCTATCACCC<br>AGACTGGAATGCACTGGCATGATCTCAGCTCACTGCAACCTCTGCCCTCCGGGTTCCAGA<br>GGTTCTCTGCCCTCAGCTCCCGAGTAGCTGGACTTCAGGCGTGCAGCTAATT<br>[A, T]<br>CTTTTTTTTTTTTTTAATTGAGACAAAGTCTCACTCTATCACCCAGACTGGAATGCA<br>GTGGCATGATCTCAGCTCACTGCAACCTCTGCCCTCCGGGTTCCAGAGGTTCTCTGCC<br>CAGCCTCCCGAGTAGCTGGACTTCAGGCTGCGCCAGTAATTGTTATTTAGTA<br>GAGACAGCTTTTACCATGTTGGCTGGCTGCTGCAACTCCGATCTCAAGCAATCTG<br>CCTGCCCTGGTCTCTAAAGTGCCTGGATTACAGGCATAAGCCACGATGCCCTTTGT |
| 20717 |  | ATGCACCAACTTGCTGTTAGTCAGTCACTGAAGCTTCTGCACTGGCTTCCCTCCCTCCA<br>GGCTCCAGGGGATGCTGAGCTCTCTTGACACTCAGAGGAGGACATTGCACTGACT<br>GGACTTAGAGCAGGTGAGCTGAGGGAAAGGGCTGTGAGGGTGGAGCAGGGCAAGAGGG<br>GAAGGATGGGTCGCTGTCAAATACAAGGCCTTCACTCAGCTGTCACCTCCAGGCCAG<br>AGCAGTCACATTCAAGGCCACAAAGATTGTGGTCACTTTGTTTCTTTCTTT<br>[T, C]<br>CTTTTTTTTTTTTAATTGAGACAAAGTCTCACTCTATCACCCAGACTGGAATGCA<br>GTGGCATGATCTCAGCTCACTGCAACCTCTGCCCTCCGGGTTCCAGAGGTTCTCTGCC<br>CAGCCTCCCGAGTAGCTGGACTTCAGGCTGCGCCAGTAATTGTTATTTAGTA<br>GAGACAGCTTTTACCATGTTGGCTGGCTGCTGCAACTCCGATCTCAAGCAATCTG<br>CCTGCCCTGGTCTCTAAAGTGCCTGGATTACAGGCATAAGCCACGATGCCCTTTGT   |
| 20894 |  | GGGGAGGGATGGGTCGCTGTCAAATACAAGGCCTTCACTCAGCTGTCACCTCCAGCC<br>CAGAGCAGTCACATTCAAGGCCACAAAGATTGTGGTCACTTTGTTTCTTTCT<br>TTTCTTTTTTTTTAATTGAGACAAAGTCTCACTCTATCACCCAGACTGGAA<br>TGCACTGGCATGATCTCAGCTCACTGCAACCTCTGCCCTCCGGGTTCCAGAGGTTCTCT<br>GCCTCAGCTCCCGAGTAGCTGGACTTCAGGCTGCGCCAGTAATTGTTATTT<br>[A, G]<br>GTAGAGACAGCTTTTACCATGTTGGCTGGGCTGGCTCGAACCTCCGATCTCAAGCAAT<br>CTGCCCTGCCCTGGCTCTAAAGTGCCTGGATTACAGGCATAAGCCACGATGCCCTGG<br>TGTGTTCTTCTCACTCCCTGAAAGGCATCGTGGGAGAGGGTAGTCAGTGGACCA<br>AGTCCTAGAGAACAGTATCTATTCTTCTCAACACATCACCCACGTGACCTGAGC<br>AAGCCACACATCACCCCTGGGCCAGTTTATCATCTGAAATTAGGGAAACATAGGT   |
| 21787 |  | GGGTGCAATGGTTCACACCTGTAATCCAGCACTTGGGAGGCTGAGGTGGCGGACAC<br>CTGAGGTCAAGGAGTTGAGACAGCCTGGCAACATGGGAAACCCGCTCTACTAAAA<br>GCACAAAATTAGCCAGCGTAGTGGTCACTGCCCTGAGTCCCAGCTACTCGGAAGCTG<br>AGGCATGAGAATCACTGAACCTGGAGGAGATGTTGCACTGAGGCCAGATCGTGC<br>TGCACTCCAGCTGGGTACAGAGCTAGACTGTCACAAACAAACAAACAAAC<br>[-, A, C]<br>TAAAAGATATGTGGATATGAGGGATCACCACCCATAGGGCCCTGGATTAAACACCC<br>CCACCAATGCCCTGAATTAAAAGAAACAGATGACTAGGTTGGAGAAATCTGGTTGG<br>GTCTATGAGAAGTAGTGTCTCTTGTGCCCTTCCCAATTCTTTGACATTGAGCTCC<br>ATGGTGCTGTCATGCCCTCTCACAGTGTGATGGCAGGTGGAGACAGATTAGAAAAG<br>AGCTGGAGCCACAGAGATTGGCAGACTGATTGCGCTCTTGGAAATCTCCAGCACA  |
| 22264 |  | CTCCATGGTGCTGTCATGCCCTCACAGTGTGATGGCAGGTGGACAGATTAGAAA   |

## FIGURE 3Q

ATAGAGCTGGAGCCACAGAGATTTGGCAGACTGATTCGGTGCCTCTTGGAAATCTCCAG  
 CACATTCCAAAAGCCTGGATAGGACAAAATAGCTTATCAACGTGAGAAAGGACTTCAG  
 AGCTTGCTACTGCCAACCTCATTTACCCAATGAGGAAAGTGAAGCTATTAGGGGGCG  
 AGGGACACGTGGAAGGTACACAGCACACAGGAGGTGATTACATGTAGATTTAGCACC  
 [T, C]  
 GCTCCTGCCACGCTGGACTGGTTACCTCTAGGCTGACCCCTGCCCTCCCCGTTCACA  
 CACACTCTGCACACACACACACACACACACAGGTGCTTGTCTGCCAGG  
 GGTTCTAGGGTACCTCTGGTTGCAGCCACTGTGACCCCAACTGGCTAACCTCTC  
 TCCCCCTCCACTTCCCTGTGGTCTGCAGGAGATTCAGCTGGGAGCATGTTACAGC  
 AGAGACCCCTGACCGCCCTGTGAGGGCTCAGCCAGTCCTCATGTCCTACAGAGTGCCTG  
 22338 ACAGAGATTTGGCAGACTGATTCGGTGCCTCTTGGAAATCTCCAGCACATTCCAAAAG  
 CCTGGATAGGACAAAATAGCTTATCAACGTGAGAAAGGACTTCAGAGCTTGCTACTGC  
 CAACCCCTCATTTACCCAATGAGGAAAGTGAAGCTATTAGGGGGCGAGGGACACGTGGAA  
 GGTACACAGCACACAGGAGGTGATTACATGTAGATTTAGCCTGCTGCCACGC  
 TGGACTGGTTACCTCTAGGCTGACCCCTGCCCTCCCCGTTCACACACACTCTGCAC  
 [-, C, A]  
 CACACACACACACACACACACACAGGTGCTTGTCTGGCCAGGGGTTCTAGGGTCA  
 CCTCTGGTTGCAGCCACTGTGACCCCAACTGGCTAACCTCTCTCCCCCTCCACTT  
 CCTCCTGTGGTCTGCAGGAGATTCAGCTGGGAGCATGTTACAGCAGAGACCTGACCG  
 CCCTGTAGGGCTCAGCCAGTCCTCATGTCCTACAGAGTGCCTGGACTGGACTTC  
 CATAAAGGATGAGCCTGGGTACAGGGGTTGTCGGCGAGGAAAGTGCATCCCCCAGA  
 23363 CAGGGACCATGTGCTCCACACCCAGGAGTCTAGGCCTGGTAACTATGCGCCCCCGT  
 CCATCATCCCCAAGGCTGCCAACACCACACTGCTGTAGCAAGCACATCAGACTCTAGC  
 CTGGACAGTGGCCAGGACCGTCGAGACCAAGAGCTACCTCCCCGGGACAGCCCACTA  
 AGGTTCTGCCCTAGCCTCTGAAACATCACTGCCCTCAGAGGCTGCTCCCTCCCCGTGA  
 GGCTGGCTAGAAACCCAAAGAGGGGATGGTAGCTGGCAGAATCATCTGGCATCTAG  
 [T, C]  
 AATAGATACCAGTTATCTGCACAAAACCTTGGAAATTCTCTTGACCCAGAGACTC  
 AGAGGGGAAGAGGGTGTAGTACCAACACAGGAAACGGATGGGACCTGGGCCAGACA  
 GTCCCCCTTGACCCCAAGGCCATCAGGAAATGCCCTTGGTAATCTGCCCTATC  
 CTTCTTACCTGGCAAAGAGCCAATCATGTTAACTCTTCTTACAGCCTGTGGCCAGA  
 GACACAATGGGTCTCTGTAGGCAAAGGTGGAAGTCCTCCAGGGATCCGCTACATCCC  
 23688 AAACTTTGGGAAATTCTCTTGCACCCAGAGACTCAGAGGGGAAGAGGGTGTAGTACC  
 AACACAGGGAAAACGGATGGGACCTGGGCCAGACAGTCCCCCTGACCCAGGGCCAT  
 CAGGGAAATGCCCTCTGGTAATCTGCCCTATCTTACCTGGCAAAGAGCAA  
 TCATGTTAACTCTTCTTACAGCCTGTGGCCAGAGACACAATGGGGTCTTCTGTAGG  
 CAAAGGTGGAAGTCCTCCAGGGATCCGCTACATCCCCTACTGATGCAGATGTGGAAAG  
 [G, A]  
 GGCTGATCCAGATTGGGCTTCTGCACAGGAAGACTCTTAAACACCTTAGGACCTCAG  
 GCCATCTTCTCTATGAAGATGAAAATAGGGGTTAAGTTCCATATGTACAAGGAGGTA  
 TTGAGAGGAACCCCTACTGTTGACTTGAAAATAATAGGTTCCATGTGAAAGTGTGTTGTA  
 AAATTTCAGTGGAAATGACAGAAAATCTCTGGCTCTCATCACTGCTTTCTCAAGCT  
 TCTTCAGCTTAACAACCCCTTCCCTAACAGGTTGGGCTGGCCAGCCTAGGAAAACATCC  
 24210 TCACTGCTTTCTCAAGCTTCAAGCTTAACAACCCCTTCCCTAACAGGTTGGCCTGGC  
 CCAGCCTAGGAAAACATCCCATTCTAACCTCAGCCAGACCTGCGTTGTGCTGTG  
 GTTGAGTGTAGCTGGTCAAGCTAACAGTCTTCTAGAGTTAAAGGAGGGGGTGTGGCAA  
 GAGCCAACACATTCTGGCCAGGAGCATTGCTTCTGTGAAATTCAATTGCACTCTGG  
 CTGCCAATGGAACTCAAAACTTGGAAAGGCGAAGGACAATGTTATCTGGATTACCGTGC  
 [A, C]  
 CAGCACCCGAAGTGCCTAAATTCCAGGAGGACAAGAGCCTTAGCCAATGACAACCTCACTCT  
 CCCCTACTCCACCTCTTCCAAGTCCAGCTCAGGCCAGGAGGTGGGAGAAGGTCAAGA  
 GCCTCAGGAATTCCAAGTCAAGTCTTGAACCAAGTATCTAGATCCCCTGAGGAC  
 TTGATGAAGTGTCTTAACCCCCAAGTAATCTTAAACCCCCAGACCCCTCAGAACTG  
 AAGGAGATTGTTGACCCAGTGACCTGGAGTTGAGGCTCAGGGAGAGATCTGCCACATGTC